

거리두기 시행 장기화에 따른 유동인구 분석

컴퓨터공학부
201702495 김수연

CONTENTS

- I. 연구 개요
- II. 데이터셋
- III. 데이터 전처리
- IV. 데이터 분석
- V. 시각화
- VI. 분석 결과

I. 연구 개요

배 경

지난 7월부터 10월까지 거리두기 4단계 시행이 3개월 이상 지속됐지만 여전히 많은 수의 확진자가 발생했다. 이는 거리두기 장기화에 따른 피로 누적 및 느슨해진 경각심으로 인해 거리두기 시행 목적과 반하여 유행인구가 증가한 것에 원인이 있을 것이라고 생각한다.

연구 목 적

서울시 거리두기 시행 시점별 유행인구를 분석하여 거리두기 시행 초기~장기화에 따른 유행인구 추이를 파악한다.

가 설

단계 격상 초기에만 유행인구가 감소하고 장기화 됨에 따라 격상 이전의 수준으로 증가할 것이다.

II. 데이터셋

1. 데이터 수집 기간 정의

3차 대유행이 시작되기 바로 전인 2020년 10월부터, 거리두기 4단계 시행이 장기화된 2021년 9월까지
(2021년 10월도 거리두기 4단계가 시행되었지만 워드코로나가 대두화되면서 분석에 영향이 있을것이라 판단)

2. 데이터셋 수집

서울시 공공데이터 활용

- 서울시 지하철 승하차 인원 정보
- 스마트서울 도시데이터 센서(S-DoT) 유동인구 측정 정보
- 한강공원 주차장 월별 이용 현황

II. 데이터셋

서울시 지하철 승하차 인원 정보

- 날짜(idx), 노선명, 역명, 승차총승객수, 하차총승객수, 등록일자, ''
- 결측치 없음
- 날짜, 하차총승객수 컬럼 분석에 사용
(수도권 지역에서의 유입을 고려, 승차승객 보다 하차승객 중에서
서울에 실질적으로 방문할 목적의 사람들이 많을 것이라고 판단)

	노선명	역명	승차총승객수	하차총승객수	등록일자	
20201001	2호선	시청	2118	1605	20201004	NaN
20201001	1호선	청량리(서울시립대입구)	6567	6637	20201004	NaN
20201001	1호선	제기동	3450	3186	20201004	NaN
20201001	1호선	신설동	3386	3300	20201004	NaN
20201001	1호선	동대문	6187	6468	20201004	NaN
...
20210930	9호선	신논현	28562	28302	20211003	NaN
20210930	9호선2~3단계	연주	9859	10044	20211003	NaN
20210930	9호선2~3단계	선정릉	9308	10135	20211003	NaN
20210930	9호선2~3단계	삼성중앙	6389	6552	20211003	NaN
20210930	9호선2~3단계	봉은사	14768	15040	20211003	NaN

218868 rows × 6 columns

II. 데이터셋

S-DOT 도시데이터 센서 유동인구 측정 정보

- 기관 명, 송신 서버 번호, 데이터 번호, 모델명, 시리얼, 서버타입, 사이트명, 방문자수, 날짜, 등록일자
- 결측치 없음
- 방문자수, 날짜 컬럼 분석에 사용

	기관 명	송신 서버 번호	데이터 번호	모델명	시리얼	서버타입	사이트명	방문자수	날짜	등록일자
0	서울시	32	1	SDOT001	2992	00Original	2992	200	202010010000	2020-10-01 00:21:39
1	서울시	32	1	SDOT001	2993	00Original	2993	186	202010010000	2020-10-01 00:21:40
2	서울시	32	1	SDOT001	2994	00Original	2994	159	202010010000	2020-10-01 00:21:40
3	서울시	32	1	SDOT001	2995	00Original	2995	141	202010010000	2020-10-01 00:21:40
4	서울시	32	1	SDOT001	2996	00Original	2996	130	202010010000	2020-10-01 00:21:41
...
27248	서울시	32	1	SDOT001	4030	00Original	4030	1	202109282350	2021-09-28 23:57:23
27249	서울시	32	1	SDOT001	4007	00Original	4007	5	202109282350	2021-09-28 23:57:28
27250	서울시	32	1	SDOT001	4040	00Original	4040	14	202109282350	2021-09-28 23:57:43
27251	서울시	32	1	SDOT001	4042	00Original	4042	0	202109282350	2021-09-28 23:57:47
27252	서울시	32	1	SDOT001	4025	00Original	4025	3	202109282350	2021-09-28 23:57:52

4025460 rows × 10 columns

II. 데이터셋

한강공원 주차장 월별 이용 현황

- 지구별, 주차장명, 주차대수, 이용시간, 날짜 컬럼으로 구성
- 결측치 없음
- 날짜, 주차대수 컬럼 분석에 사용

	지구별	주차장명	주차대수	이용시간	날짜
0	PLT-005	독섬4주차장	708	68086	2021/10
1	PLT-007	양화2주차장	834	69036	2021/10
2	PLT-007	양화2주차장	1019	83361	2021/10
3	PLT-003	망원23주차장	11415	1112191	2021/10
4	PLT-007	양화1주차장	3928	390477	2021/10
...
28491	PLT-008	여의도5주차장	2805	860432	2020/01
28492	PLT-009	반포23주차장	23590	1200177	2020/01
28493	PLT-003	망원23주차장	1382	231375	2020/01
28494	PLT-005	독섬1주차장	1529	379777	2020/01
28495	PLT-004	이촌2주차장	3002	144936	2020/01

28496 rows × 5 columns

III. 데이터 전처리 서울시 지하철 승하차 인원 정보

```
1 # 사용하지 않는 컬럼 제거
2 subway = s.drop(['노선명', '역명', '승차총승객수', '등록일자', ''], axis=1)
3 subway.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 218868 entries, 20201001 to 20210930
Data columns (total 1 columns):
#   Column  Non-Null Count  Dtype
---  -
0   하차총승객수  218868 non-null     int64
dtypes: int64(1)
memory usage: 3.3 MB
```

```
1 # 동일 일자의 데이터 합산하여 데이터프레임 재구성
2 date_s = [20201001, 20201101, 20201201, 20210101, 20210201, 20210301, 2021
3 date_e = [20201031, 20201130, 20201231, 20210131, 20210228, 20210331, 2021
4
5 df = {}
6 for s, e in zip(date_s, date_e):
7     for d in range(s, e+1):
8         date = str(d)[:4] + '-' + str(d)[4:6] + '-' + str(d)[6:]
9         df[date] = sum(subway.loc[d, '하차총승객수'])
10 subway = pd.DataFrame(df.items(), columns=['날짜', '총승객수'])
11 subway['날짜'] = pd.to_datetime(subway['날짜'])
12 subway.set_index('날짜', inplace=True)
13 subway
```

총승객수

날짜

2020-10-01	1775583
2020-10-02	2441986
2020-10-03	2993972
2020-10-04	2701215
2020-10-05	6159619

... ..

2021-09-26	3074753
2021-09-27	6261902
2021-09-28	6273005
2021-09-29	6053994
2021-09-30	6379926

365 rows × 1 columns

III. 데이터 전처리 S-DOT 도시데이터 센서 유동인구 측정 정보

```
1 # 사용하지 않는 컬럼 제거
2 visited = walk.drop(['기관명', '송신 서버 번호', '데이터 번호', '모델명', '시리
3 visited = visited[['날짜', '방문자수']]
4 visited.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 4025460 entries, 0 to 27252
Data columns (total 2 columns):
#   Column  Dtype
---  -
0   날짜      object
1   방문자수  object
dtypes: object(2)
memory usage: 92.1+ MB
```

```
1 visited.set_index('날짜', inplace=True)
2 visited
```

방문자수	
날짜	
202010010000	200
202010010000	186
202010010000	159
202010010000	141
202010010000	130
...	
202109282350	1
202109282350	5
202109282350	14
202109282350	0
202109282350	3

4025460 rows × 1 columns

(* '월-총 방문자수' 형식의 데이터로 처리하는 과정에서 에러 발생하여 수정중)

III. 데이터 전처리 한강공원 주차장 월별 이용 현황

```
1 # 사용하지 않는 컬럼 제거
2 parking = p.drop(['지구별', '주차장명', '이용시간'], axis=1)
3 parking = parking[['날짜', '주차대수']]
```

```
1 parking.set_index('날짜', inplace=True)
2 parking.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 28496 entries, 2021/10 to 2020/01
Data columns (total 1 columns):
#   Column  Non-Null Count  Dtype
---  -
0   주차대수    28496 non-null   int64
dtypes: int64(1)
memory usage: 445.2+ KB
```

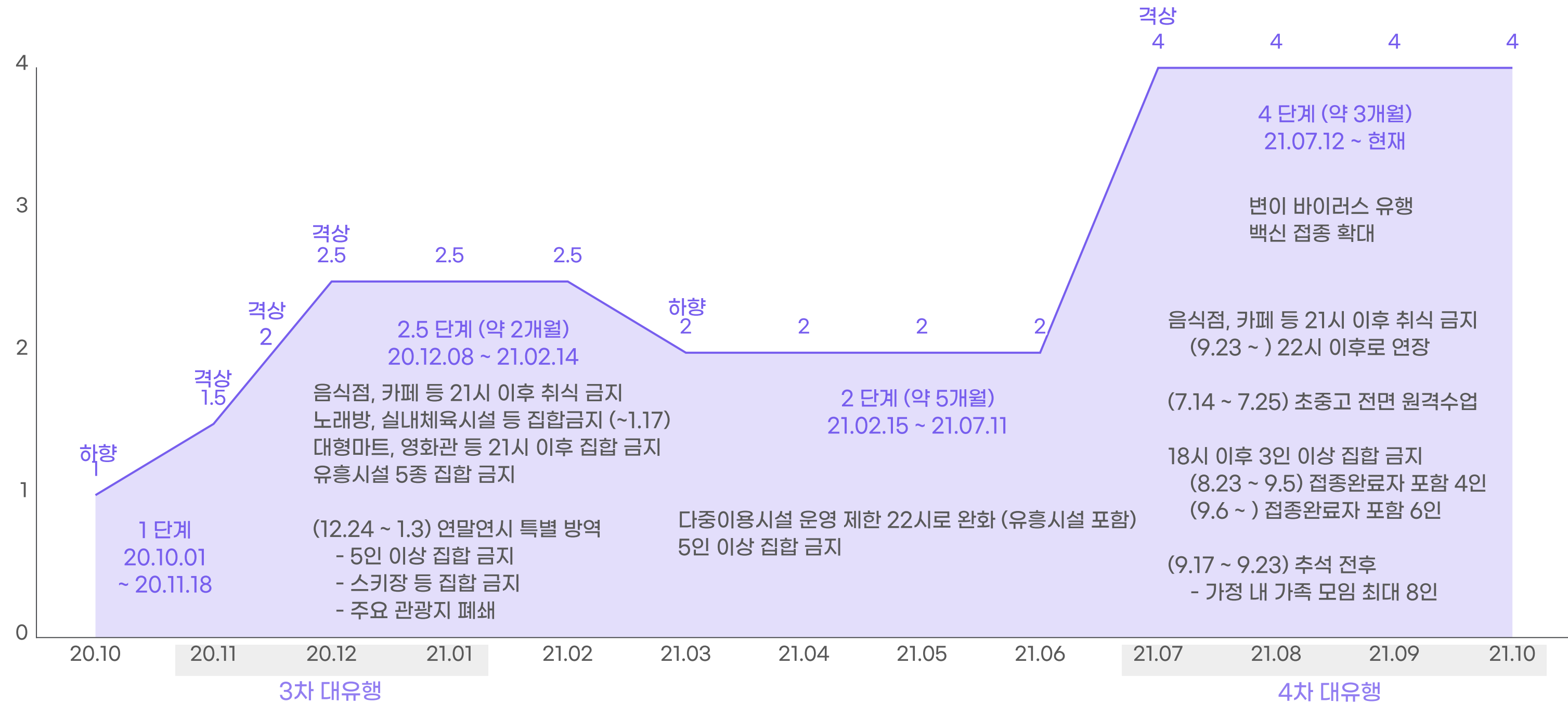
```
1 # 동일 월의 데이터 합산하여 데이터프레임 재구성
2 date = ['2020/10', '2020/11', '2020/12', '2021/01', '2021/02', '2021/03', '2021/04', '2021/05', '2021/06', '2021/07', '2021/08', '2021/09']
3
4 df = {}
5 for d in date:
6     df[d] = sum(parking.loc[d, '주차대수'])
7 parking = pd.DataFrame(df.items(), columns=['날짜', '총주차대수'])
8 parking['날짜'] = pd.to_datetime(parking['날짜'])
9 parking.set_index('날짜', inplace=True)
10 parking
```

총주차대수

날짜

2020-10-01	38565126
2020-11-01	31146444
2020-12-01	32647266
2021-01-01	29960280
2021-02-01	33179814
2021-03-01	39042486
2021-04-01	46669716
2021-05-01	45453366
2021-06-01	48143214
2021-07-01	39471030
2021-08-01	33424837
2021-09-01	29625119

IV. 데이터 분석 2020.10 ~ 2021.09 서울시 거리두기 시행 주요 타임라인



IV. 데이터 분석 데이터셋 분할 구간 정의

기간	거리두기 단계	구간 특징	분석 고려사항
20.10.01 ~ 20.11.18 (7주)	1	- 단계 하향 - 3차 대유행 시작 전	거리두기 강도가 낮을 때의 유동인구수 파악
20.11.19 ~ 21.01.03 (약 6주)	1.5 -> 2 -> 2.5 (+ 연말 특별 방역)	- 지속적 단계 격상 - 고강도 거리두기 시행 - 일 확진자 최초 1천명 돌파	지속적으로 단계 격상시의 유동인구 추이
21.01.04 ~ 21.02.14 (6주)	2.5	- 고강도 거리두기 장기화	장기화에 따른 유동인구 추이
21.02.15 ~ 21.02.28 (2주)	2	- 국민적 피로감을 고려해 단계 하향	단계 하향시 유동인구 추이
21.03.01 ~ 21.07.11 (19주)	2	- 약 4개월간 2, 3주 간격으로 거리두기 연장 발표 - 4차 대유행 시작 전	장기화에 따른 유동인구 추이
21.07.12 ~ 21.07.25 (2주)	4	- 단계 격상 - 변이 바이러스 유행	고강도 격상 초반의 유동인구 추이
21.07.26 ~ 21.10.03 (10주)	4	- 고강도 거리두기 장기화 - 8월 이후 백신 접종 가속화	격상 초반 2주와 유동인구 추이 비교