

데이터 베이스(01)

DB응용프로그램 작성 보고서 제출하기

1770117 호선우

1. 사용한 데이터

서울특별시 코로나19 백신 예방접종 현황.csv

<http://data.seoul.go.kr/dataList/OA-20914/S/1/datasetView.do>

접종일, 백신 1차, 2차 접종누적수 및 접종률 등을 알 수 있다.

A	B	C	D	E	F	G	H
접종일	접종대상지	당일접종	일차접종누적	일차접종률	당일2차접종	2차접종누적	2차접종률
#####	2195709	950	912517	41.6	0	325079	14.8
#####	2196518	33401	911559	41.5	3580	325079	14.8
#####	2188641	105089	877710	40.1	17084	321446	14.7
#####	2175217	122034	772541	35.5	13061	304336	14
#####	1129593	10780	650432	57.6	16175	291118	25.8
#####	1123393	7748	639483	56.9	21010	274749	24.5
#####	1124159	7748	631695	56.2	21010	253528	22.6
#####	1123388	7688	623942	55.5	21524	231821	20.6
#####	1122778	0	623941	55.6	0	231820	20.6
#####	1121838	2478	621463	55.4	4100	227493	20.3
#####	1123023	2271	619017	55.1	28449	198625	17.7
#####	1121981	2205	616705	55	22755	175445	15.6
#####	1120058	14	616699	55.1	3724	171720	15.3
#####	1122759	2983	613630	54.7	16649	154898	13.8
#####	1119118	3197	610403	54.5	13074	141769	12.7
#####	1118166	0	610338	54.6	0	141769	12.7
#####	1118491	1021	609382	54.1	3796	137973	11.2
#####	1113909	3827	604185	54.2	12751	125220	11.2
#####	1116796	3391	600548	53.8	12435	112781	10.1
#####	1114297	1617	598266	53.7	13970	98810	8.9
#####	1113580	2796	594975	53.4	13030	85775	7.7
#####	1113434	3131	591040	53.1	12457	73295	6.6
#####	1112095	0	589744	53	0	73003	6.6
#####	1119900	4044	592073	52.9	2082	70926	6.3
#####	1108162	11230	574583	51.9	11486	59707	5.4
#####	1106344	11021	563044	50.9	11715	47940	4.3
#####	1104738	1788	550895	50.8	1650	46700	4.1

서울특별시 코로나19 확진자 발생동향.csv

<http://data.seoul.go.kr/dataList/OA-20461/S/1/datasetView.do>

기준일, 코로나 누적 확진자 및 추가 확진자 수를 알 수 있다.

A	B	C	D
서울시기준	서울시확진	서울시추가확진	
#####	43917	130	
#####	43787	160	
#####	43627	194	
#####	43433	214	
#####	43219	218	
#####	43001	262	
#####	42739	138	
#####	42601	139	
#####	42462	181	
#####	42281	208	
#####	42073	172	
#####	41901	242	
#####	41659	249	
#####	41410	172	
#####	41238	195	

2. 응용 프로그램 및 환경

사용한 DBMS : MySQL Workbench 8.0

사용도구 : PyMySQL

3. 분석 내용

- ① "서울특별시 코로나19 확진자 발생동향"을 이용하여 2020-02-05부터 2021-05-31까지 코로나 누적 확진자 수를 matplotlib 라이브러리를 이용하여 그래프로 보여준다. (한달 단위)

➔ 코로나 증가 추세를 보기 위함

- ② "서울특별시 코로나19 확진자 발생동향"을 이용하여 2020-02-05부터 2021-05-31까지 코로나 추가 확진자 수를 matplotlib 라이브러리를 이용하여 하나의 그래프에 보여준다. (한달 단위)

➔ 코로나 증가 추세를 보기 위함

- ③ "서울특별시 코로나19 확진자 발생동향" 과 "서울특별시 코로나19 백신 예방접종 현황"을 이용하여 2021-04-21부터 2021-05-30까지 코로나 추가 확진자 및 1차 및 2차 백신 접종 누계를 matplotlib 라이브러리를 이용하여 그래프로 보여준다. (2일 단위)

➔ 백신 증가 추세에 따른 코로나 추가 확진자 수를 보기 위함

- ④ "서울특별시 코로나19 확진자 발생동향"을 이용하여 2020-02-05부터 2021-05-31까지 추가 확진자 500명이 넘었던 날짜를 검색한다.

- ⑤ "서울특별시 코로나19 확진자 발생동향" 과 "서울특별시 코로나19 백신 예방접종 현황"을 이용하여 2021-04-21부터 2021-05-30까지 코로나 추가 확진자 및 1차 및 2차 백신 접종률을 검색한다.

4. DB 구축

PyMySQL을 이용한 DB생성

```
import pymysql

db = pymysql.connect(host='localhost', port=3306, user='root', password='1234',
charset='utf8')

cursor = db.cursor()
```

```
sql = 'CREATE DATABASE coronaVaccine'

cursor.execute(sql)
db.commit()
```

corona, vaccine 테이블 생성

```
db = pymysql.connect(host='localhost', port=3306, user='root', password='1234',
db='coronaVaccine', charset='utf8')
cursor = db.cursor()

sql = '''
CREATE TABLE corona (
서울시기준일 VARCHAR(30) NOT NULL PRIMARY KEY,
서울시확진자 INT(20) NOT NULL,
서울시추가확진 INT(20) NOT NULL
)
'''

cursor.execute(sql)

db.commit()
db.close()

db = pymysql.connect(host='localhost', port=3306, user='root', password='1234',
db='coronaVaccine', charset='utf8')
cursor = db.cursor()

sql = '''
CREATE TABLE vaccine (
접종일 VARCHAR(30) NOT NULL,
접종대상자 INT(20) NOT NULL,
당일일차접종자수 INT(20) NOT NULL,
일차접종누계 INT(20) NOT NULL,
일차접종률 FLOAT(5, 1),
당일이차접종자수 INT(20) NOT NULL,
이차접종누계 INT(20) NOT NULL,
이차접종률 FLOAT(5, 1),
FOREIGN KEY(접종일)
REFERENCES corona(서울시기준일)
)
'''

cursor.execute(sql)

db.commit()
db.close()
```

csv파일 적재

```
import pandas as pd

db = pymysql.connect(host='localhost', port=3306, user='root', password='1234',
db='coronaVaccine', charset='utf8')
cursor = db.cursor()
```

```

data = pd.read_csv('서울특별시 코로나19 확진자 발생동향.csv', encoding='cp949')

sql = "INSERT INTO corona (서울시기준일, 서울시확진자, 서울시추가확진) VALUES
(%s, %s, %s)"
for i in range(len(data)):
    cursor.execute(sql, (str(data['서울시기준일'][i]), int(data['서울시확진자'][i]),
int(data['서울시추가확진'][i])))

db.commit()
db.close()

db = pymysql.connect(host='localhost', port=3306, user='root', password='1234',
db='coronaVaccine', charset='utf8')
cursor = db.cursor()

data = pd.read_csv('서울특별시 코로나19 백신 예방접종 현황.csv', encoding='cp949')

sql = "INSERT INTO vaccine (접종일, 접종대상자, 당일일차접종자수, 일차접종누계,
일차접종률, 당일이차접종자수, 이차접종누계, 이차접종률) VALUES
(%s, %s, %s, %s, %s, %s, %s, %s)"

for i in range(len(data)):
    cursor.execute(sql, (str(data['접종일'][i]), int(data['접종대상자'][i]),
int(data['당일일차접종자수'][i]), int(data['일차접종누계'][i]),
float(data['일차접종률'][i]), int(data['당일이차접종자수'][i]),
int(data['이차접종누계'][i]), float(data['이차접종률'][i])))

db.commit()
db.close()

```

DB Table

corona

	서울시기 준일	서울시확 진자	서울시추가 확진
▶	2020-02-05	8	0
	2020-02-06	11	0
	2020-02-07	10	0
	2020-02-08	10	0
	2020-02-09	10	0
	2020-02-10	10	0
	2020-02-11	10	0
	2020-02-12	10	0
	2020-02-13	10	0
	2020-02-14	10	0
	2020-02-15	10	0
	2020-02-16	11	0
	2020-02-17	12	0
	2020-02-18	12	0
	2020-02-19	13	0
	2020-02-20	15	0
	2020-02-21	18	0
	2020-02-22	24	0
	2020-02-23	25	0

vaccine

	접종일	접종대상자	당일일차접종 자수	일차접종 누계	일차접 종률	당일이차접종 자수	이차접종 누계	이차접 종률
▶	2021-05-30	2195709	950	912517	41.6	0	325079	14.8
	2021-05-29	2196518	33401	911559	41.5	3580	325079	14.8
	2021-05-28	2188641	105089	877710	40.1	17084	321446	14.7
	2021-05-27	2175217	122034	772541	35.5	13061	304336	14.0
	2021-05-26	1129593	10780	650432	57.6	16175	291118	25.8
	2021-05-25	1123393	7748	639483	56.9	21010	274749	24.5
	2021-05-24	1124159	7748	631695	56.2	21010	253528	22.6
	2021-05-23	1123388	7688	623942	55.5	21524	231821	20.6
	2021-05-22	1122778	0	623941	55.6	0	231820	20.6
	2021-05-21	1121838	2478	621463	55.4	4100	227493	20.3
	2021-05-20	1123023	2271	619017	55.1	28449	198625	17.7
	2021-05-19	1121981	2205	616705	55.0	22755	175445	15.6
	2021-05-18	1120058	14	616699	55.1	3724	171720	15.3
	2021-05-17	1122759	2983	613630	54.7	16649	154898	13.8
	2021-05-16	1119118	3197	610403	54.5	13074	141769	12.7
	2021-05-15	1118166	0	610338	54.6	0	141769	12.7
	2021-05-14	1118491	1021	609382	54.1	3796	137973	11.2
	2021-05-13	1113909	3827	604185	54.2	12751	125220	11.2

5. SQL문을 사용하여 데이터 검색 및 결과 그래프 생성

- ① “서울특별시 코로나19 확진자 발생동향”을 이용하여 2020-02-05부터 2021-05-31까지 코로나 누적 확진자 수를 matplotlib 라이브러리를 이용하여 그래프로 보여준다. (한달 단위)

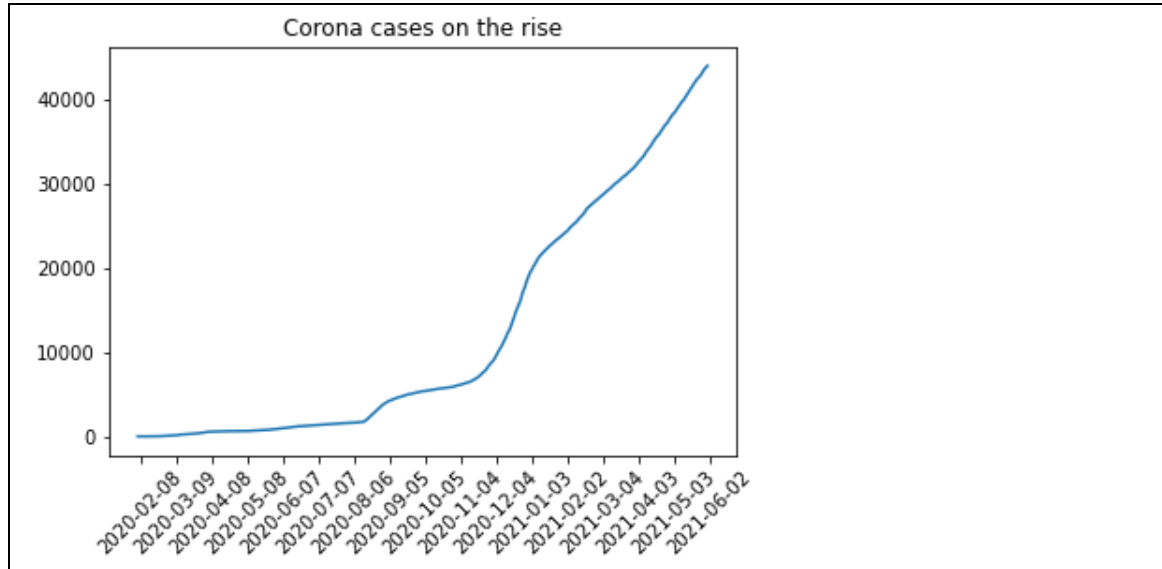
사용한 SQL문

SELECT * FROM corona

```
db = pymysql.connect(host='localhost', port=3306, user='root', password='1234',
db='coronaVaccine', charset='utf8')
SQL = '''SELECT * FROM corona;'''
corona = pd.read_sql(SQL, db)

import matplotlib.pyplot as plt
import pandas as pd
import matplotlib.ticker as ticker
import numpy as np

x = pd.date_range(corona.서울시기준일[0], corona.서울시기준일[len(corona) - 1])
y = corona.서울시확진자
plt.plot(x, y)
plt.xticks(rotation=45)
ax=plt.gca()
ax.xaxis.set_major_locator(ticker.MultipleLocator(30))
plt.title('Corona cases on the rise')
```

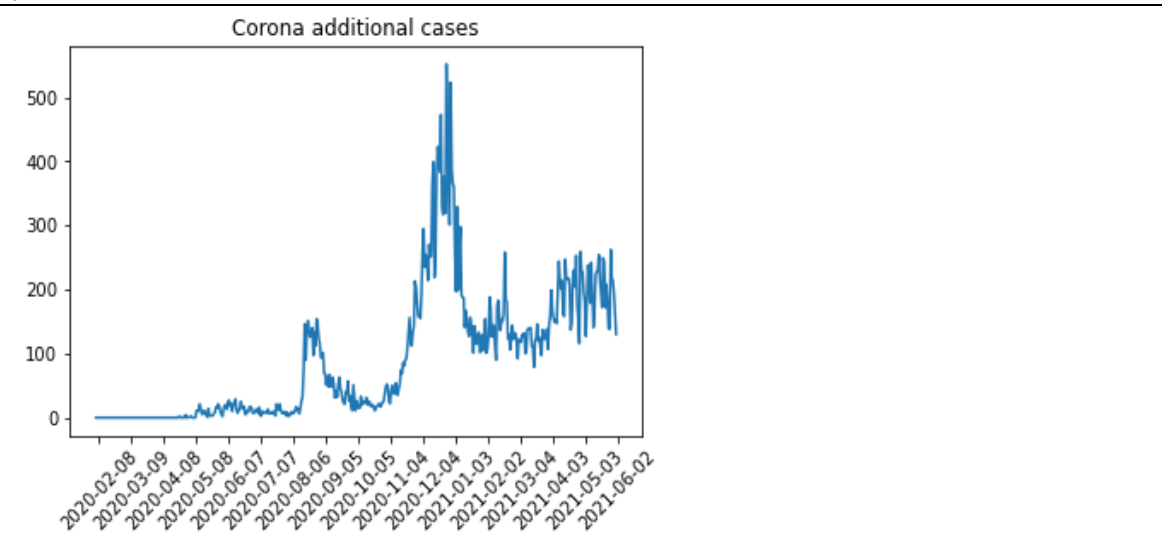


- ② “서울특별시 코로나19 확진자 발생동향”을 이용하여 2020-02-05부터 2021-05-31까지 코로나 추가 확진자 수를 matplotlib 라이브러리를 이용하여 하나의 그래프에 보여준다. (한달 단위)

사용한 SQL문

```
SELECT * FROM corona
```

```
import matplotlib.pyplot as plt
import pandas as pd
import matplotlib.ticker as ticker
import numpy as np
x = pd.date_range(corona.서울시기준일[0], corona.서울시기준일[len(corona) - 1])
y = corona.서울시추가확진
plt.plot(x, y)
plt.xticks(rotation=45)
ax=plt.gca()
ax.xaxis.set_major_locator(ticker.MultipleLocator(30))
plt.title('Corona additional cases')
```



- ③ “서울특별시 코로나19 확진자 발생동향” 과 “서울특별시 코로나19 백신 예방접종 현황”을 이용하여 2021-04-21부터 2021-05-30까지 코로나 추가 확진자 및 1차 및 2차 백신 접종 누계를 matplotlib 라이브러리를 이용하여 그래프로 보여준다. (2일 단위)

사용한 SQL문

```
SELECT 접종일 FROM vaccine
```

```
SELECT 일차접종누계 FROM vaccine ORDER BY 일차접종누계 ASC
```

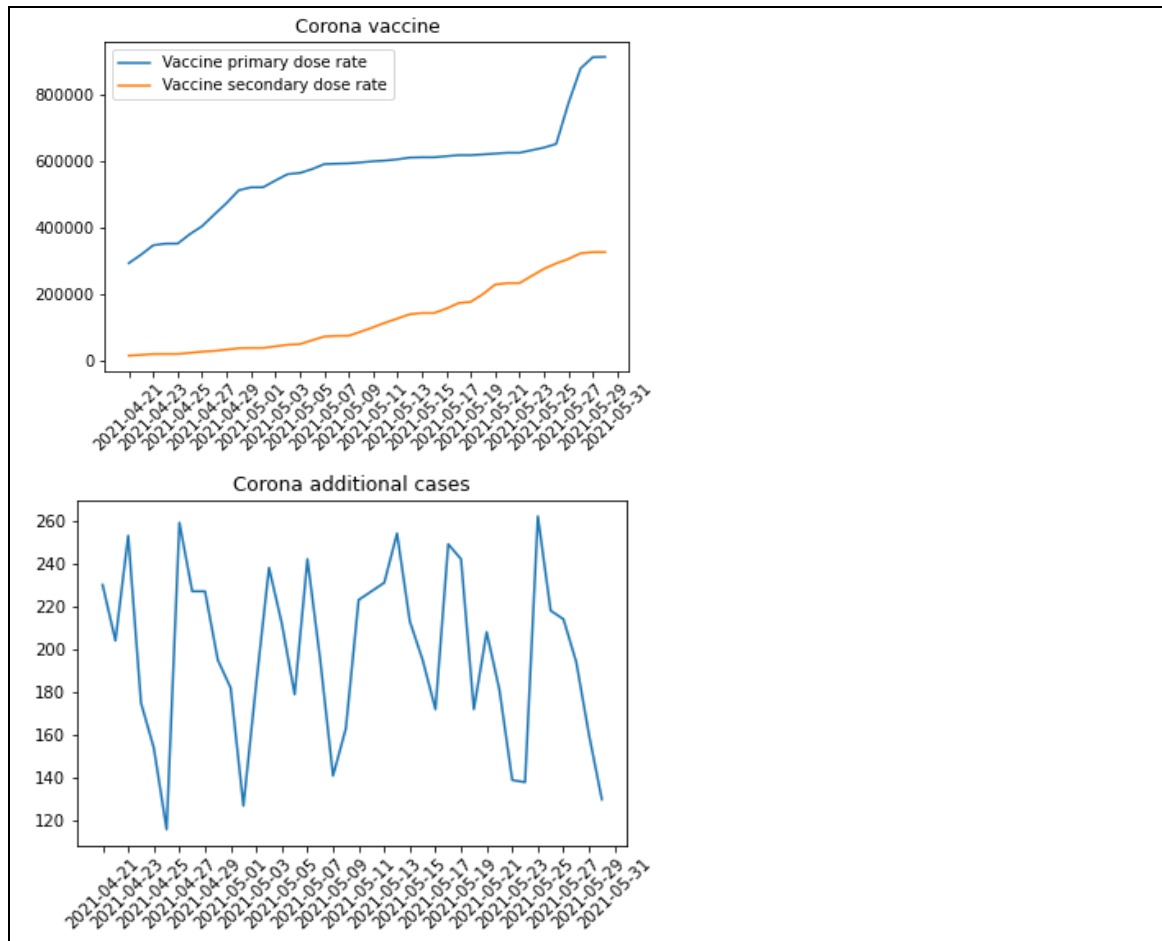
```
SELECT 이차접종누계 FROM vaccine ORDER BY 이차접종누계 ASC
```

```
db = pymysql.connect(host='localhost', port=3306, user='root', password='1234',
db='coronaVaccine', charset='utf8')
SQL = '''SELECT 접종일 FROM vaccine;'''
vaccine_date = pd.read_sql(SQL, db)
SQL = '''SELECT 일차접종누계 FROM vaccine ORDER BY 일차접종누계 ASC;'''
vaccine_1 = pd.read_sql(SQL, db)
SQL = '''SELECT 이차접종누계 FROM vaccine ORDER BY 이차접종누계 ASC;'''
vaccine_2 = pd.read_sql(SQL, db)

import matplotlib.pyplot as plt
import pandas as pd
import matplotlib.ticker as ticker
import numpy as np

x = pd.date_range(vaccine_date.접종일[0], vaccine_date.접종일[len(vaccine_date) - 1])
y = vaccine_1
z = vaccine_2
plt.figure(1)
plt.plot(x, y)
plt.plot(x, z)
plt.xticks(rotation=45)
ax=plt.gca()
ax.xaxis.set_major_locator(ticker.MultipleLocator(2))
plt.title('Corona vaccine')
plt.legend(["Vaccine primary dose rate", "Vaccine secondary dose rate"])

x = pd.date_range(vaccine_date.접종일[0], vaccine_date.접종일[len(vaccine_date) - 1])
k = corona.서울시추가확진.loc[442:481]
plt.figure(2)
plt.plot(x, k)
plt.xticks(rotation=45)
ax=plt.gca()
ax.xaxis.set_major_locator(ticker.MultipleLocator(2))
plt.title('Corona additional cases')
```



- ④ “서울특별시 코로나19 확진자 발생동향”을 이용하여 2020-02-05부터 2021-05-31까지 추가 확진자 500명이 넘었던 날짜를 검색한다.

사용한 SQL문

SELECT 서울시기준일 FROM corona WHERE 서울시추가확진 > 500

```
db = pymysql.connect(host='localhost', port=3306, user='root', password='1234',
db='coronaVaccine', charset='utf8')
SQL = '''SELECT 서울시기준일 FROM corona WHERE 서울시추가확진 > 500;'''
print("서울시 코로나 추가 확진자 500명이 넘은 날짜")
corona_max500 = pd.read_sql(SQL, db)
corona_max500
```

서울시기준일

0	2020-12-25
1	2020-12-29

- ⑤ “서울특별시 코로나19 확진자 발생동향” 과 “서울특별시 코로나19 백신 예방접종 현황”을 이용하여 2021-04-21부터 2021-05-30까지 코로나 추가 확진자 및 1차 및 2차 백신 접종률을 검색한다.

사용한 SQL문

SELECT 서울시기준일, 서울시추가확진, 일차접종률, 이차접종률 FROM corona, vaccine WHERE corona.서울시기준일 = vaccine.접종일

```
db = pymysql.connect(host='localhost', port=3306, user='root', password='1234',
db='coronaVaccine', charset='utf8')
SQL = '''SELECT 서울시기준일, 서울시추가확진, 일차접종률, 이차접종률 FROM corona,
vaccine WHERE corona.서울시기준일 = vaccine.접종일;'''
print("서울시 코로나 추가 확진자와 백신 접종률")
corona_vac = pd.read_sql(SQL, db)
corona_vac
```

	서울시기준일	서울시추가확진	일차접종률	이차접종률
0	2021-05-30	160	41.6	14.8
1	2021-05-29	194	41.5	14.8
2	2021-05-28	214	40.1	14.7
3	2021-05-27	218	35.5	14.0
4	2021-05-26	262	57.6	25.8
5	2021-05-25	138	56.9	24.5
6	2021-05-24	139	56.2	22.6
7	2021-05-23	181	55.5	20.6
8	2021-05-22	208	55.6	20.6
9	2021-05-21	172	55.4	20.3
10	2021-05-20	242	55.1	17.7
11	2021-05-19	249	55.0	15.6
12	2021-05-18	172	55.1	15.3
13	2021-05-17	195	54.7	13.8
14	2021-05-16	213	54.5	12.7
15	2021-05-15	254	54.6	12.7
16	2021-05-14	231	54.1	11.2
17	2021-05-13	227	54.2	11.2
18	2021-05-12	223	53.8	10.1
19	2021-05-11	163	53.7	8.9
20	2021-05-10	141	53.4	7.7
21	2021-05-09	195	53.1	6.6
22	2021-05-08	242	53.0	6.6
23	2021-05-07	179	52.9	6.3
24	2021-05-06	212	51.9	5.4
25	2021-05-05	238	50.9	4.3
26	2021-05-04	184	50.6	4.1

27	2021-05-03	127	49.0	3.8
28	2021-05-02	182	47.2	3.3
29	2021-05-01	195	47.2	3.3
30	2021-04-30	227	46.6	3.3
31	2021-04-29	227	43.6	2.9
32	2021-04-28	259	40.2	2.5
33	2021-04-27	116	37.3	2.4
34	2021-04-26	154	35.1	2.0
35	2021-04-25	175	34.2	1.8
36	2021-04-24	253	34.2	1.8
37	2021-04-23	204	33.6	1.8
38	2021-04-22	230	31.0	1.6
39	2021-04-21	218	28.6	1.3