**#서울시 코로나19 확진자 코드**

**import** pandas **as** pd

**import** numpy **as** np

**from** statsmodels.tsa.arima\_model **import** ARIMA

**import** statsmodels.api **as** sm

**from** fbprophet **import** Prophet

**import** matplotlib.pyplot **as** plt

**import** seaborn **as** sns

file\_name **=** f"seoul\_covid2.csv"

file\_name

covid\_19 **=** pd**.**read\_csv(file\_name)

covid\_19**.**shape

covid\_19 **=** covid\_19**.**sort\_values(["연번"], ascending **=** **False**)

covid\_19**.**head()

covid\_19**.**tail()

**import** matplotlib.pyplot **as** plt

plt**.**rc("font", family**=**"AppleGothic")

plt**.**rc("axes", unicode\_minus **=** **False**)

plt**.**style**.**use("fivethirtyeight")

**from** IPython.display **import** set\_matplotlib\_formats

set\_matplotlib\_formats("retina")

covid\_19["확진일"]**.**head()

covid\_19["확진일자"] **=** pd**.**to\_datetime("2020"**+** "-" **+** covid\_19["확진일"]**.**str**.**replace(".","-"))

covid\_19["확진일자"]**.**head()

covid\_19["월"] **=** covid\_19["확진일자"]**.**dt**.**month

covid\_19["주"] **=** covid\_19["확진일자"]**.**dt**.**isocalendar()**.**week

*#서울시 월별 확진자 수 추이*

covid\_19["확진일자"]**.**value\_counts()**.**sort\_index()**.**plot(

figsize**=**(15, 4), rot **=** 100, grid **=** **True**)

plt**.**axhline(50, linestyle**=**":", color **=** "red")

*#일자별 확진자 수 선 프로그래프 #odc\_count => oneday\_count*

covid\_19["월일"] **=** covid\_19["확진일자"]**.**astype(str)**.**map(**lambda** x : x[**-**5:])

odc\_count **=** covid\_19["월일"]**.**value\_counts()**.**sort\_index()

b **=** odc\_count**.**plot(figsize**=**(15, 4), rot **=** 30)

*#cc => case\_count*

**for** i **in** range(len(odc\_count)):

cc **=** odc\_count**.**iloc[i]

**if** cc **>** 250:

b**.**text(x **=** i, y **=** odc\_count**.**iloc[i], s **=** odc\_count**.**iloc[i])

*#일별 확진자 수 요약 #odc\_count => oneday\_count*

odc\_count**.**describe()

*#확진자 수가 가장 많았던 날 #odc\_count => oneday\_count*

odc\_count[odc\_count **==** odc\_count**.**max()]

*#확진자가 가장 많았던 날 발생 이력 찾기*

covid\_19[covid\_19["월일"] **==** "12-03"]**.**head()

b **=** odc\_count**.**plot**.**bar(figsize **=** (100, 4), rot **=** 100)

b**.**axhline(100, linestyle**=**":")

*#cc => case\_count*

**for** i **in** range(len(odc\_count)):

cc **=** odc\_count**.**iloc[i]

**if** cc **>** 10:

b**.**text(x **=** i, y **=** odc\_count**.**iloc[i], s **=** odc\_count**.**iloc[i])

*#cc => case\_count*

odc\_count\_last **=** odc\_count[**-**50:]

b**=** odc\_count\_last**.**plot**.**bar(figsize**=**(24,4))

b**.**axhline(odc\_count\_last**.**median(), linestyle**=**":", color**=**"red")

**for** i **in** range(len(odc\_count\_last)):

cc **=** odc\_count\_last**.**iloc[i]

**if** cc **>** 200:

b**.**text(x **=** i, y **=** odc\_count\_last**.**iloc[i], s **=** odc\_count\_last**.**iloc[i])

*#월별 확진자 수에 대한 빈도 수 #mc => month case*

mc **=** covid\_19["월"]**.**value\_counts()**.**sort\_index()

b **=** mc**.**plot**.**bar(rot **=** 20)

**for** i **in** range(len(mc)):

mc\_i **=** mc**.**iloc[i]

**if** mc\_i **>** 0:

b**.**text(x **=** i, y **=** mc\_i, s **=** mc\_i)

*#wc => week case*

wc **=** covid\_19["주"]**.**value\_counts()**.**sort\_index()

*#2020-01-01부터 주 단위 빈도 수*

wc**.**plot**.**bar(rot **=** 0, figsize**=**(15, 4))

*#월과 주 합치기 #mwc => month week case*

mwc **=** covid\_19**.**groupby(["월", "주"])["연번"]**.**count()

*#월과 주 합친 그래프*

mwc**.**plot**.**bar(figsize**=**(15, 4), rot **=** 100)

*#모든 날짜를 행에 만들기*

firtst **=** covid\_19**.**iloc[**-**1]["확진일자"]

last **=** covid\_19**.**iloc[0]["확진일자"]

firtst, last

*#pd.date\_range를 통해 시작하는 날짜부터 끝나는 날짜까지의 DatetimeIndex를 만들고 days라는 변수에 저장*

days **=** pd**.**date\_range(firtst, last)

days[:5]

*#days 변수의 값 ="년도 월 일" 컬럼 이름을 갖는 dateframe 생성*

days **=** pd**.**DataFrame({"확진일자": days})

*#dc => day case*

dc **=** covid\_19["확진일자"]**.**value\_counts()

dc **=** dc**.**to\_frame()

dc**.**columns **=** ["확진자 수"]

*#확진자가 없는 날도 일자에 표현할 수 있도록 전체 일자와 확진 데이터를 merge로 합치기 #ad => all day*

ad **=** days**.**merge(dc, left\_on**=**"확진일자", right\_on**=**dc**.**index, how**=**"left")

ad **=** ad**.**rename(columns**=**{"연번":"확진자 수"})

*#누적 확진자 수*

ad["누적 확진"] **=** ad["확진자 수"]**.**fillna(0)**.**cumsum()

*#년도를 제외하고 월-일 -> "일자" 컬럼 만들기*

ad["일자"] **=** ad["확진일자"]**.**astype(str)**.**map(**lambda** x: x[**-**5:])

*#"확진자수", "누적확진자 수" 컬럼을 갖는 데이터프레임 생성*

accumulate\_day **=** ad**.**set\_index("일자")

accumulate\_day **=** accumulate\_day[["확진자 수","누적 확진"]]

*#확진 월과 요일 구하기*

ad["확진 월"] **=** ad["확진일자"]**.**dt**.**month

ad["확진 요일"] **=** ad["확진일자"]**.**dt**.**dayofweek

*#월별, 요일별 확진자 수 groupby 구하기*

ad\_week **=** ad**.**groupby(

["확진 월", "확진 요일"])["확진자 수"]**.**sum()**.**unstack()**.**astype(int)

ad\_week

*#확진 요일 숫자로 표현된 것을 문자로 변경*

oneweek**=** "월 화 수 목 금 토 일"

oneweek **=** oneweek**.**split()

*#컬럼의 이름을 요일명으로 변경*

ad\_week**.**columns **=** oneweek

ad\_week**.**style**.**background\_gradient(cmap**=**"Greens")

*#거주지역 확진자 수 빈도수 구하기*

area\_gu **=** covid\_19["지역"]**.**value\_counts()

area\_gu**.**sort\_values()**.**plot**.**barh(figsize**=**(15, 12), grid **=** **True**)