#데이터 로드 SeoulFloating.csv

df\_서울\_유동인구 = pd.read\_csv("/gdrive/My Drive/2020\_ai\_project1/data/SeoulFloating.csv", encoding = "UTF-8")

# 설정 조건에 따라, 데이터셋 정제

# "province" 행 삭제

df\_서울\_유동인구.drop("province", axis = 1, inplace = True)

#"birth\_year"이 70인 행 삭제

drop\_70s = df\_서울\_유동인구[df\_서울\_유동인구["birth\_year"] == 70].index

df\_서울\_유동인구.drop(drop\_70s, axis = 0, inplace = True)

#시간대 18시 ~ 23시

df\_서울\_유동인구 = df\_서울\_유동인구[(df\_서울\_유동인구["hour"] >= 18) & (df\_서울\_유동인구["hour"] <= 23)]

#서울 내 유흥주점 개수 상위 3개 구 선택

condition = (df\_서울\_유동인구["city"] == "Gangnam-gu") | (df\_서울\_유동인구["city"] == 'Gwangjin-gu') | (df\_서울\_유동인구["city"] == 'Mapo-gu')

df\_서울\_유동인구 = df\_서울\_유동인구[condition)

#코로나 발생 전 금토일 데이터 추출

condition\_전 = (df\_서울\_유동인구["date"] == "2020-01-17") | (df\_서울\_유동인구["date"] == "2020-01-18") | (df\_서울\_유동인구["date"] == "2020-01-19")

df\_서울\_유동인구\_전 = df\_서울\_유동인구[condition\_전]

#코로나 발생 전 시간대별 유동인구 합계 계산

sum\_by\_date\_hour\_city\_전 = df\_서울\_유동인구\_전["fp\_num"].groupby([df\_서울\_유동인구\_전["date"], df\_서울\_유동인구\_전["hour"], df\_서울\_유동인구\_전["city"]]).sum()

sum\_by\_date\_hour\_city\_전 = sum\_by\_date\_hour\_city\_전.unstack().reset\_index() #멀티인덱스 없애주기

#문자열로 통일

sum\_by\_date\_hour\_city\_전["hour"] = sum\_by\_date\_hour\_city\_전["hour"].astype(str)

#새로운 컬럼 생성

for i, date in enumerate(sum\_by\_date\_hour\_city\_전["date"]):

if date == '2020-01-17':

sum\_by\_date\_hour\_city\_전.loc[i,"date\_hour"] = "금요일 " + sum\_by\_date\_hour\_city\_전.loc[i,"hour"]

elif date == '2020-01-18':

sum\_by\_date\_hour\_city\_전.loc[i,"date\_hour"] = "토요일 " + sum\_by\_date\_hour\_city\_전.loc[i,"hour"]

else:

sum\_by\_date\_hour\_city\_전.loc[i,"date\_hour"] = "일요일 " + sum\_by\_date\_hour\_city\_전.loc[i,"hour"]

#필요 없는 컬럼 삭제

sum\_by\_date\_hour\_city\_전.drop("date", axis = 1, inplace = True)

sum\_by\_date\_hour\_city\_전.drop("hour", axis = 1, inplace = True)

sum\_by\_date\_hour\_city\_전 = sum\_by\_date\_hour\_city\_전.rename( #인덱스 변경

{

'Gangnam-gu': 'Gangnam-gu\_전',

'Gwangjin-gu': 'Gwangjin-gu\_전',

'Mapo-gu' : 'Mapo-gu\_전',

'date\_hour' : "date\_hour"

}, axis = "columns"

)

#코로나 발생 후 - 고강도 사회적거리두기 기간 중 금토일 데이터에도 동일한 과정 반복(코드 생략)

#전후 데이터셋 병합

sum\_by\_date\_hour\_city = pd.merge(sum\_by\_date\_hour\_city\_전, sum\_by\_date\_hour\_city\_후, on = "date\_hour")

#구별 전후 유동인구 수 그래프화 ex. 강남구

fig\_유동인구\_강남 = px.line(sum\_by\_date\_hour\_city, x = "date\_hour", y = ["Gangnam-gu\_전", "Gangnam-gu\_후"], title = '강남구 금토일 주요 시간대 유동인구 수 변화')

py.iplot(fig\_유동인구\_강남)