

- 1 Exercise. 서울시 구별 CCTV 현황 분석하기
- 2
- 3 -시험시간 : 1시간
- 4
- 5 Task1. CCTV 현황과 인구 현황 데이터 구하기
- 6
- 7 1. Google에서 '서울시 자치구 연도별 cctv 설치 현황'이라고 검색.
- 8 2. 검색결과페이지에서 '[FILE]서울시 자치구 연도별 CCTV 설치 현황...'
page(<https://opengov.seoul.go.kr/data/2813904>)로 들어간다.
- 9 3. [내용 바로가기] 섹션에서 '[FILE]서울시 자치구 연도별 CCTV 설치 현황' 우측의 [내용보기] link를 click한다.
- 10 4. <http://data.seoul.go.kr/dataList/datasetView.do?infId=OA-2734&srvType=C&serviceKind=1>
- 11 5. 페이지 하단의 Chart | File 중 File을 클릭한다.
- 12 6. '서울시 자치구 연도별 CCTV 설치 현황.xlsx'파일을 다운로드한다.
- 13 7. 다시 Google 사이트에서 '서울시 서울통계'로 검색한다.
- 14 8. 첫 검색 page인 '서울통계서비스 - 서울 열린데이터광장 - 서울시'로 들어간다.
- 15 9. 인구 > 주민등록인구 > 주민등록인구(구별)을 클릭한다.
- 16 10.
- 17 <https://data.seoul.go.kr/dataList/datasetView.do?serviceKind=2&infId=419&srvType=S&stcSrl=419>
- 18 11. 페이지 하단에서 [XLS]를 선택하여 'report.xls'파일을 다운로드받는다.
- 19
- 20 Task2. pandas를 이용해서 CCTV와 인구 현황 데이터 파악하기
- 21 1. 변수명 CCTV_Seoul에 Task1에서 다운로드받았던 '서울시 자치구 연도별 CCTV 설치 현황.xlsx'을
DataFrame으로 변환하여 저장한다.
- 22 2. 이 데이터를 '소계'별로 오름차순 소팅한다.
- 23 3. CCTV 최근증가율을 알아내기 위해 2014년부터 2017년까지 최근 4년간 CCTV수를 더하고, 2013년 이전
CCTV수로 나눠서 최근 4년간 CCTV 증가율을 계산한다.
- 24 4. 서울시 인구현황을 파악하기 위해, Task1에서 다운로드받았던 '주민등록인구(구별)'의 데이터를 pop_Seoul 변수에
DataFrame으로 저장한다.
- 25 5. header는 3번째부터, 그리고 칼럼은 B,D,G,J,N만 저장한다.
- 26 6. 칼럼의 이름은 '구별', '인구수', '한국인', '외국인', '고령자'로 설정한다.
- 27 7. 0번째 행은 각 항목의 합계를 나타내기 때문에 우리에게 필요하지 않아서 0번째 행을 삭제한다.
- 28 8. '구별' 열의 unique를 조사한다.
- 29 9. 혹시 NaN행이 있으면 삭제한다.
- 30 10. 각 구별 전체 인구를 이용해서 구별 '외국인 비율'과 '고령자비율'을 계산하자.
- 31 11. 각 구별 외국인 비율은 각 외국인합계를 각 구별 인구수로 나눠서 계산하고, 각 구별 고령자 비율은 각 구별 고
령자수를 각 구별 인구수로 나눠서 계산한다.
- 32 12. 이렇게 만들어진 DataFrame을 인구수로 내림차순, 외국인 수별로 내림차순, 외국인 비율과 고령자 비율로 각
각 내림차순 소팅한다.
- 33 13. 다시 외국인 수별로 내림차순 소팅한다.
- 34
- 35
- 36 Task3. CCTV 데이터와 인구 현황 데이터를 합치고 분석하기
- 37 1. 위에서 만든 2개의 DataFrame을 pandas의 merge()를 이용해서 합친다. 이때 공통 컬럼인 '구별'기준으로 합
친다.
- 38 2. 더 이상 의미없는 칼럼인 '2013년 이전', '2014년', '2015년', '2016년', '2017년' 칼럼을 삭제한다.
- 39 3. 지금까지의 DataFrame의 칼럼의 순서는 '구별', '소계', '최근증가율', '인구수', '한국인', '외국인', '고령자', '외
국인비율', '고령인비율'이다.
- 40 4. 합쳐진 DataFrame의 index를 '구별'로 지정한다. (hint : set_index())
- 41 5. NumPy의 상관계수를 구하는 함수인 corrcoef()를 이용해서 '고령자비율'과 '소계'를, '외국인비율'과 '소계'를,
마지막으로 '인구수'와 '소계'를 각각 상관계수를 조사한다.
- 42 6. 상관계수는 절대값이 클수록 두 데이터가 관계가 있다고 본다.
- 43 7. 절대값이 0.1 이하면 거의 무시, 0.3이하면 약한 상관관계, 0.7이상이면 뚜렷한 상관관계가 있다고 본다.