|  |  |
| --- | --- |
| 교육 제목 | **판다스 데이터 살펴보기, 통계함수, 내장 그래프 활용** |
| 교육 일시 | 2021-09-30 |
| 교육 장소 | C06 |
| **교육 내용** | |
| 오전 | 데이터 살펴보기    파일을 가져와 내용 불러오기  데이터 내용 미리보기  데이터 요약정보 확인하기    **Shape 속성을 이용하여 행과 열 개수를 튜플 로 보여주며,**  **데이터 크기를 확인 하려면 df.shape 입력하여 사용**  **데이터프레임 기술 통계 정보 요약하기**  **: describe() 메소드를 사용하여 산술(숫자) 데이터를 갖는 열에 대한 주요 기술 통계 정보를 확인 할 수 있다**   * **평균, 표준편차, 최댓값, 최솟값, 중간값 등**     기술통계 정보 확인하기 실습  **데이터 개수 확인하기**    **Info()메소드 화면에 count() 메소드를 사용하여 각 열이름과**  **데이터 개수를 시리즈 객체로 출력함**  판다스 자료형 비교   |  |  | | --- | --- | | 판다스 | 파이썬 | | Int64 | Int | | Float64 | Float | | object | String | | Datetime64, timedelta64 | 없음 |   **통계함수 적용**  평균값 : df.mean() -> 전체 컬럼 값의 평균  중간값 : df.median() -> 전체 컬럼 값의 중간값  최대값 : df.max()-> 전체 컬럼 값의 최대값  최소값 : df.min() -> 전체 컬럼 값의 최소값  표준편차 : df.std() -> 전체 컬럼 값의 표준편차 값  산관계수 : df.corr() -> 전체 컬럼 값의 상관계수 값    간단한 통계함수 사용해보기        **판다스 내장 그래프 도구 활용**    Datafile에 저장 되 있는 남북한발전전력량에서 데이터 값 가져와서 출력하기    내장 그래프 plot 사용하여 그래프로 표시하기    내장 그래프 자주 사용하는 그래프들로 출력  시각화 도구  : matplotlib -> 대체로 주식, 시세 변동 등에 사용됨    Matplotlib 사용하여 데이터 시각화 해보기  데이터 전처리 작업 예제    중간점검 문제  1.부산에서 다른 지역으로 전출한 데이터만 추출해서 df\_busan으로 저장  2.불필요한 컬럼 제거, 인덱스는 전입지별 -> 전입지로 바꾸어서 설정    점검 문제 코드 & 실행 결과  문제코드 활용 matplotlib 사용해보기 1    문제코드 활용 matplotlib 사용해보기 2 |
| 오후 | 다양한 그래프 활용해보기    Marker 사용해서 점 표시해보기      **여러가지 그래프 같이 출력하기**  면적 그래프 사용해보기      **막대 그래프 사용해보기**    산점도 그래프 활용해보기      Seaborn 라이브러리 : 고급 그래프 도구   * Seaborn + matplotlib의 기능 스타일을 확장한 시각화 고급 버전       간단한 seaborn 라이브러리 사용해보기  Folium 라이브러리 : 지도 활용  지도 위에 시각화 할 때 유용한 도구, 세계지도를 포함한 여러가지 지도지원        Folium 라이브러리 사용하여 세계지도 출력  오늘의 활용문제  문제  - iris데이터를 sns 모듈에서 load하여 종별로 잎의 넓이와 길이를  산점도와 선 그래프로 표시하세요  - legend(종) 와 제목("붓꽃의 잎의 길이와 넒이, 종의 관계")도 표시  **풀이**    **%load graph.py 불러와서**    **코드와 출력 화면** |