|  |  |
| --- | --- |
| 교육 제목 | **Pandas 데이터 살펴보기, 시각화** |
| 교육 일시 | 2021년 09월 30일 |
| 교육 장소 | YGL-C6 |
| **교육 내용** | |
| 오전 | 1. 데이터 내용 미리보기   객체.head() 메소드 앞부분 미리보기  객체.tail() 메소드 뒷부분 미리보기  텍스트이(가) 표시된 사진  자동 생성된 설명   1. 데이터 요약 정보 확인하기   객체.shape() 데이터프레임의 크기 확인 메소드  객체.info() 데이터프레임의 기본 정보(클래스 유형, 행 인덱스 구성, 열 이름 종류 개수 등등)   1. 데이터프레임의 기술 통계 정보 요약   객체.describe() 산술 데이터를 갖는 열에 대한 주요 기술 통계 정보(평균, 표준편차, 최댓값,최솟값, 중간값 등 요약해서 객체로 반환  loc 메소드를 사용해서 원하는 값만 확인 가능     1. 데이터 개수 확인   객체.count() 각 열이 가지고 있는 원소 개수 확인  객체.value\_counts() 특정 열이 가지고 있는 고유값 확인   1. 통계 함수 적용   mean, median, std, max, min등등 통계값을 구할 수 있다.  사용방법은 객체 뒤에 .mean(), .median(), .max()등으로 사용 가능하다.  텍스트, 영수증, 스크린샷이(가) 표시된 사진  자동 생성된 설명테이블이(가) 표시된 사진  자동 생성된 설명  mean() : 산술 데이터를 갖는 모든 열의 평균값  median(): 산술 데이터를 가진 열의 중간 값 계산  max() :산술 데이터를 가진 열의 최대값  min() : 산술 데이터를 가진 열의 최솟값  std() : 산술 데이터를 가진 열의 표준 편차  corr(): 산술 데이터를 가진 두 열의 상관계수 <= 분석때 가장 중요   1. Pandas 내장 그래프 도구 활용   시각화는 효과적인 정보 전달 방식, 판다스 내장함수에 시각화 지원   1. 선 그래프      1. 박스 그래프   텍스트이(가) 표시된 사진  자동 생성된 설명     1. 산점도   텍스트이(가) 표시된 사진  자동 생성된 설명     * 하지만 시각화를 할 때 더 다양한 그래프를 그린다. 이를 위한 라이브러리로는 matplotlib와 seaborn이 존재한다. |
| 오후 | 1. matplotlib   matplotlib은 파이선 표준 시각화 도구라고 부를 수 있을 정도로 2d 평면 그래프에 관한 다양한 포맷과 기능을 지원한다. 객체 지향 프로그래밍 되어 있는 패키지로 그래프를 세세하게 꾸밀 수 있다.   * 시도별 전출입 인구수를 시각화 하기 위해 데이터를 불러 온 뒤 전처리 * 텍스트이(가) 표시된 사진    자동 생성된 설명 * 텍스트이(가) 표시된 사진    자동 생성된 설명 * 텍스트이(가) 표시된 사진    자동 생성된 설명   plt.title()그래프의 제목을 추가  plt.xlabel() x축의 이름 추가  plt.ylabel() y축의 이름 추가  plt.legend() 그래프의 범주 추가  plt.xticks() x축의 글자 모양 지정 (rotation =45 는 45도 돌리는 것으로 글자가 겹치는걸 방지)  plt.show() 만들어진 그래프 출력   * 1. 그래프 꾸미기     figure()함수로 그래프 틀의 사이즈 설정  color keyword를 통해 그래프별 색깔 변경  marker keyword를 통해 마킹 모양 변경   * 1. 그래프 분할   plt.figure()를 통해 빈 캠버스를 생성하고 add\_subplot()을 통해 그래프를 추가 시켜 준다. 하지만 이러한 방법은 여러개의 그래프를 일일히 다 그려야 하기 때문에 불편하다.  그래서 fig,axs = plt.subplots(nrow ,ncol)를 이용하여서 여러개의 ax를 한꺼번에 생성 시켜 주었다. 이를 이용하면 나중에 그래프를 꾸밀 때 axs가 시퀀스 형으로 나오기 때문에 반복문을 활용하여 쉽게 꾸며 줄 수 있다.    pandas에서 제공하는 plot메소드에서 kind = ‘area’ 키워드를 통해서 면적 그래프를 구할 수 있다. 데이터를 누적하는 방법 (stacked = True), 누적하지 않는 방법(stacked = False)를 사용 하여 나타낼 수 있다.    kind 키워드에 bar를 넣으면 막대 그래프를 나타낼 수 있고 barh를 이용하면 수평막대그래프를 나타낼 수 있다. 두가지 방식으로 plot하기 위해 데이터를 전처리한 후 살펴 보았다.    보조축 활용하기   * 지금까지는 그래프를 그릴 때 y축을 한개만 사용 하였지만 보조 축을 추가하여 2개의 축을 갖는 그래프를 그릴 수 있다.   텍스트이(가) 표시된 사진  자동 생성된 설명    위 결과 처럼 두가지 그래프를 나타낼 때 한개의 x축과 두개의 y축을 이용 하여 나타낼 수 있다.  그 이외에 여러가지 plot을 하는 방법이 있지만 따로 실습 진행 하지 않음  scatter(산점도) ,hist(히스토그램), boxplot(박스플롯) pie(파이차트)...   1. Seaborn 라이브러리   seaborn은 matplotlib의 기능과 스타일을 확장한 파이썬 시각화 도구의 고급버전이다.      실습을 진행하면서 seaborn의 버전이 책과 맞지 않아 warning이 발생 하였다. 이는 pip install –upgrade seaborn명령어를 통해 해결  seaborn에는 여러가지 그래프를 지원해준다. 그 중 히트맵이 가장 많이 쓰임 (집가서 실습할 것)   1. folium 라이브러리   folium 라이브러리는 지도 위에 시각화 할 때 유용한 도구이다 .  텍스트이(가) 표시된 사진  자동 생성된 설명  지도이(가) 표시된 사진  자동 생성된 설명  지도에 마커 표시 하기  엑셀에 저장되어 있는 서울 소재 대학교의 좌표값을 판다스를 이용해 데이터프레임으로 불러와서 해당 좌표에 있는 이름을 지도에 마킹 한다.  텍스트이(가) 표시된 사진  자동 생성된 설명  지도이(가) 표시된 사진  자동 생성된 설명 |