

빅데이터분석 & 시각화 평가자체크리스트 평가지

과 정 명	[K디지털]AI&빅데이터 분석 활용 폴스택 서비스 개발	과 목 명	빅데이터 분석결과 시각화																																																																																									
평가일자	2025-05-27	평가자	박경미																																																																																									
평가유형	평가자 체크리스트	평가단위 요소명	- 통계분석 및 시각화 - 지도분석 및 시각화하기 - 텍스트 분석 및 시각화																																																																																									
평가문항 (수행내용)	◆ 다음의 요구사항에 따라 시각화를 구현하시오.																																																																																											
	- 5장 기술통계분석과 상관관계 분석 강의 자료에 통계분석 실습자료를 참고하여 [구현1:기술통계 분석] 구현하라.																																																																																											
	[구현 1: 기술통계분석 및 상관관계 분석]																																																																																											
	1) data4폴더에 저장된 삶의 만족도.xlsx, 건강.xlsx, 경제.xlsx, 사회참여.xlsx, 교육.xlsx, 안전.xlsx, 여가.xlsx, 환경.xlsx를 판다스의 데이터프레임으로 로드한다.																																																																																											
	2)로드된 데이터 프레임을 다음과 같이 하나의 데이터프레임으로 합친다.																																																																																											
	<table><tr><th>No</th><th>시도</th><th>구군</th><th>삶의 만족도</th><th>건강</th><th>경제</th><th>사회참여</th><th>교육</th><th>안전</th><th>여가</th><th>환경</th></tr><tr><td>0</td><td>1</td><td>서울특별시 종로구</td><td>0.4437</td><td>0.9220</td><td>1.0000</td><td>0.7425</td><td>0.6839</td><td>0.7470</td><td>0.6331</td><td>0.4637</td></tr><tr><td>1</td><td>2</td><td>서울특별시 중구</td><td>0.4976</td><td>0.6742</td><td>0.9806</td><td>0.4608</td><td>0.5013</td><td>0.9320</td><td>0.6691</td><td>0.2865</td></tr><tr><td>2</td><td>3</td><td>서울특별시 용산구</td><td>0.6161</td><td>0.5898</td><td>0.6915</td><td>0.4317</td><td>0.2679</td><td>0.5537</td><td>0.2817</td><td>0.5030</td></tr><tr><td>3</td><td>4</td><td>서울특별시 성동구</td><td>0.4729</td><td>0.4794</td><td>0.6533</td><td>0.4182</td><td>0.2464</td><td>0.5347</td><td>0.3257</td><td>0.4196</td></tr><tr><td>4</td><td>5</td><td>서울특별시 광진구</td><td>0.4041</td><td>0.6373</td><td>0.4445</td><td>0.3519</td><td>0.4879</td><td>0.6072</td><td>0.3313</td><td>0.4992</td></tr></table>			No	시도	구군	삶의 만족도	건강	경제	사회참여	교육	안전	여가	환경	0	1	서울특별시 종로구	0.4437	0.9220	1.0000	0.7425	0.6839	0.7470	0.6331	0.4637	1	2	서울특별시 중구	0.4976	0.6742	0.9806	0.4608	0.5013	0.9320	0.6691	0.2865	2	3	서울특별시 용산구	0.6161	0.5898	0.6915	0.4317	0.2679	0.5537	0.2817	0.5030	3	4	서울특별시 성동구	0.4729	0.4794	0.6533	0.4182	0.2464	0.5347	0.3257	0.4196	4	5	서울특별시 광진구	0.4041	0.6373	0.4445	0.3519	0.4879	0.6072	0.3313	0.4992																							
	No	시도	구군	삶의 만족도	건강	경제	사회참여	교육	안전	여가	환경																																																																																	
	0	1	서울특별시 종로구	0.4437	0.9220	1.0000	0.7425	0.6839	0.7470	0.6331	0.4637																																																																																	
	1	2	서울특별시 중구	0.4976	0.6742	0.9806	0.4608	0.5013	0.9320	0.6691	0.2865																																																																																	
	2	3	서울특별시 용산구	0.6161	0.5898	0.6915	0.4317	0.2679	0.5537	0.2817	0.5030																																																																																	
	3	4	서울특별시 성동구	0.4729	0.4794	0.6533	0.4182	0.2464	0.5347	0.3257	0.4196																																																																																	
	4	5	서울특별시 광진구	0.4041	0.6373	0.4445	0.3519	0.4879	0.6072	0.3313	0.4992																																																																																	
	3) 데이터의 결측치 값을 제거한다. 단 삶의 만족도는 평균값으로 채워서 결측치 값을 제거하고 나머지는 null 값을 포함한 행을 제거한다.																																																																																											
	4)다음과 같이 통계요약을 수행하라																																																																																											
	<table><tr><th></th><th>No</th><th>삶의 만족도</th><th>건강</th><th>경제</th><th>사회참여</th><th>교육</th><th>안전</th><th>여가</th><th>환경</th></tr><tr><td>count</td><td>223.000000</td><td>223.000000</td><td>223.000000</td><td>223.000000</td><td>223.000000</td><td>223.000000</td><td>223.000000</td><td>223.000000</td><td>223.000000</td></tr><tr><td>mean</td><td>114.950673</td><td>0.495216</td><td>0.405873</td><td>0.392500</td><td>0.470994</td><td>0.542274</td><td>0.455517</td><td>0.463204</td><td>0.577374</td></tr><tr><td>std</td><td>65.925239</td><td>0.267121</td><td>0.181590</td><td>0.211245</td><td>0.199592</td><td>0.235800</td><td>0.190406</td><td>0.229254</td><td>0.181753</td></tr><tr><td>min</td><td>1.000000</td><td>0.005200</td><td>0.005500</td><td>0.008600</td><td>0.009400</td><td>0.013900</td><td>0.031100</td><td>0.023800</td><td>0.073300</td></tr><tr><td>25%</td><td>59.500000</td><td>0.292850</td><td>0.268300</td><td>0.230850</td><td>0.339250</td><td>0.370550</td><td>0.313850</td><td>0.278200</td><td>0.455200</td></tr><tr><td>50%</td><td>115.000000</td><td>0.487900</td><td>0.389700</td><td>0.378800</td><td>0.473200</td><td>0.563200</td><td>0.446100</td><td>0.447800</td><td>0.598700</td></tr><tr><td>75%</td><td>170.500000</td><td>0.699550</td><td>0.525100</td><td>0.513150</td><td>0.617850</td><td>0.730250</td><td>0.602500</td><td>0.647400</td><td>0.701300</td></tr><tr><td>max</td><td>229.000000</td><td>0.999000</td><td>1.000000</td><td>1.000000</td><td>1.000000</td><td>1.000000</td><td>1.000000</td><td>1.000000</td><td>1.000000</td></tr></table>				No	삶의 만족도	건강	경제	사회참여	교육	안전	여가	환경	count	223.000000	223.000000	223.000000	223.000000	223.000000	223.000000	223.000000	223.000000	223.000000	mean	114.950673	0.495216	0.405873	0.392500	0.470994	0.542274	0.455517	0.463204	0.577374	std	65.925239	0.267121	0.181590	0.211245	0.199592	0.235800	0.190406	0.229254	0.181753	min	1.000000	0.005200	0.005500	0.008600	0.009400	0.013900	0.031100	0.023800	0.073300	25%	59.500000	0.292850	0.268300	0.230850	0.339250	0.370550	0.313850	0.278200	0.455200	50%	115.000000	0.487900	0.389700	0.378800	0.473200	0.563200	0.446100	0.447800	0.598700	75%	170.500000	0.699550	0.525100	0.513150	0.617850	0.730250	0.602500	0.647400	0.701300	max	229.000000	0.999000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000
	No	삶의 만족도	건강	경제	사회참여	교육	안전	여가	환경																																																																																			
count	223.000000	223.000000	223.000000	223.000000	223.000000	223.000000	223.000000	223.000000	223.000000																																																																																			
mean	114.950673	0.495216	0.405873	0.392500	0.470994	0.542274	0.455517	0.463204	0.577374																																																																																			
std	65.925239	0.267121	0.181590	0.211245	0.199592	0.235800	0.190406	0.229254	0.181753																																																																																			
min	1.000000	0.005200	0.005500	0.008600	0.009400	0.013900	0.031100	0.023800	0.073300																																																																																			
25%	59.500000	0.292850	0.268300	0.230850	0.339250	0.370550	0.313850	0.278200	0.455200																																																																																			
50%	115.000000	0.487900	0.389700	0.378800	0.473200	0.563200	0.446100	0.447800	0.598700																																																																																			
75%	170.500000	0.699550	0.525100	0.513150	0.617850	0.730250	0.602500	0.647400	0.701300																																																																																			
max	229.000000	0.999000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000																																																																																			
5) 위의 데이터프레임에서 시도필드를 value_count를 구하라(결과 생략: 수업자료 참고)																																																																																												
6) 데이터프레임으로 시도로 그룹을 하여 삶의 만족도와 건강의 평균과 합을 구하라.																																																																																												
7) 데이터프레임으로 시도로 그룹을 하여 삶의 만족도는 평균과 중앙값을, 건강은 합과 표준편차를 구하라.																																																																																												
8) 위의 데이터프레임으로 삶의 만족도를 종속변수로 나머지 필드는 독립변수하여 회귀 분석을 수행하라																																																																																												
9) 회귀분석결과를 시각화하라.																																																																																												
10) 작성된 회귀분석모델을 시각화를 수행하라.																																																																																												
11) 위의 데이터프레임으로 필드들 사이의 상관관계 분석을 수행하고 상관관계 분석 결과로 heatmap으로 시각화하라																																																																																												

[구현2] 네이버 개발자 센터에 뉴스를 검색하여 뉴스의 discription으로 텍스트를 사용하
요 워드클라우드를 작성한다.

- 1) 네이버뉴스의 검색어는 자유롭게 선택하여 1000개의 뉴스를 가져온다.
- 2) 뉴스에서 검색어를 사용하여 해당 뉴스를 검색하여 제목(title), 본문(description), 링크(link, originallink), 작성일(pubDate)를 json 형태로 읽어서 파일로 저장한다.
- 3) 저장된 뉴스 데이터에서 본문(description)으로 형태소 분석기로 명사를 추출한다.
- 4) 추출한 명사를 출현 빈도수를 계산하고 가장 인기 있는 단어 50개를 추출한다.
- 5) 글자의 길이가 2자 이상인 것만 추출한다.
- 6) 위의 데이터를 사용하여 워드클라우드를 작성한다.

[구현3: 광역시도별 삶의 만족도 choropleth map 시각화]

- 1) 광역시도별 groupby함수를 사용하여 삶의 만족도 평균을 구한다.
- 2) 광역시도 행정구역 geojson 데이터와 광역시도별 삶의 만족도 평균 데이터로 choropleth map를 작성하라.

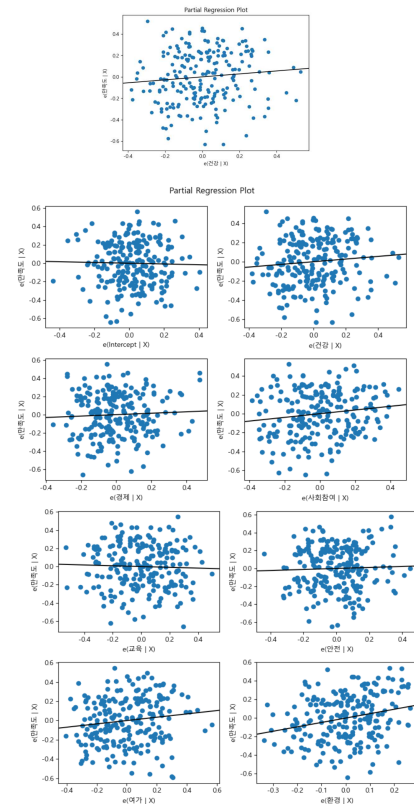
◆ 제출방법

- 1) 작성한 소스코드(.ipynb)와 결과를 캡처하여 한글답지 파일(.hwp)에 붙여서 본인이름의 폴더에 저장, 폴더를 압축하러이름.zip로 저장하여 제출함 (예: 홍길동.zip)
- 2) 제출한 상태에서 임포트한 라이브러리 및 관련 리소스들이 제대로 연동되어야 함

◆ [구현1 결과 예시]

과제물결
과
(예시)

OLS Regression Results						
Dep. Variable:	만족도			R-squared:	0.184	
Model:	OLS			Adj. R-squared:	0.158	
Method:	Least Squares			F-statistic:	6.941	
Date:	Tue, 20 May 2025			Prob (F-statistic):	1.87e-07	
Time:	12:16:20			Log-Likelihood:	1.1660	
No. Observations:	223			AIC:	13.67	
Df Residuals:	215			BIC:	40.93	
Df Model:	7					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
Intercept	-0.0407	0.118	-0.343	0.732	-0.274	0.193
건강	0.1385	0.095	1.464	0.145	-0.048	0.325
경제	0.0789	0.106	0.747	0.456	-0.129	0.287
사회참여	0.1961	0.090	2.178	0.030	0.019	0.374
교육	-0.0434	0.082	-0.529	0.598	-0.205	0.118
안전	0.0564	0.104	0.540	0.590	-0.149	0.262
여가	0.1731	0.086	2.012	0.045	0.004	0.343
환경	0.4746	0.126	3.775	0.000	0.227	0.722
Omnibus:	5.538	Durbin-Watson:		2.205		
Prob(Omnibus):	0.063	Jarque-Bera (JB):		3.994		
Skew:	-0.192	Prob(JB):		0.136		
Kurtosis:	2.469	Cond. No.		17.3		



[구현2 결과 예시]



[구현3 예시]

