

Lecture 12

Analysis and Visualization of Numerical Data

<https://pandas.pydata.org/>

1. Pandas 소개

1.1 Pandas 기초

1.2 Pandas를 활용한 UCI 자동차 연비 데이터 분석

2. Matplotlib 소개

2.1 Matplotlib를 활용한 데이터 분석

2.2 Pandas와 Matplotlib를 활용한 남북한 발전 전력량 데이터 분석

3. Seaborn 소개

4. Numpy 소개

5. Exercise

■ Pandas

- 데이터 분석을 위해 Python 프로그래밍 언어로 작성된 소프트웨어 라이브러리
 - **BSD(Berkeley Software Distribution)** 라이선스에 의해 공개된 자유 소프트웨어

Use Guide

<https://pandas.pydata.org/pandas-docs/stable/reference/>

<https://pandas.pydata.org/pandas-docs/stable/reference/series.html>

<https://pandas.pydata.org/pandas-docs/stable/reference/frame.html>

1. Pandas 소개

1.1 Pandas 기초

1.2 Pandas를
활용한 UCI
자동차 연비
데이터 분석

2. Matplotlib 소개

2.1 Matplotlib를
활용한 데이터
분석

2.2 Pandas와
Matplotlib를
활용한 남북한
발전 전력량
데이터 분석

3. Seaborn 소개

4. Numpy 소개

5. Exercise

- Pandas 데이터 구조 이해를 위한 Series 와 DataFrame 활용법 실습
 - 12_(3 page) 강의용 Pandas 기초.ipynb

<https://archive.ics.uci.edu/ml/datasets/auto+mpg>

캘리포니아 대학교 어사인(University of California, Irvine)

- UCI 자동차 연비 데이터셋은 9개의 항목으로 398개의 데이터셋으로 구성

No.	속성(attributes)		데이터 상세(범위)
1	mpg	연비	연속 값
2	cylinders	실린더 수	이산 값(예: 3, 4, 6, 8)
3	displacement	배기량	연속 값
4	horsepower	출력	연속 값
5	weight	차중	연속 값
6	acceleration	가속능력	연속 값
7	model_year	출시년도	이산 값(예: 70, 71, 80, 81)
8	origin	제조국	이산 값(예: 1(USA), 2(EU), 3(JPN))
9	name	모델명	문자열

<https://archive.ics.uci.edu/ml/datasets/auto+mpg>

캘리포니아 대학교 어사인(University of California, Irvine)

■ UCI 자동차 연비 데이터셋은 9개의 항목으로 398개의 데이터셋으로 구성

18.0,8,307.0,130.0,3504.,12.0,70,1,"chevrolet chevelle malibu"
15.0,8,350.0,165.0,3693.,11.5,70,1,"buick skylark 320"
18.0,8,318.0,150.0,3436.,11.0,70,1,"plymouth satellite"
16.0,8,304.0,150.0,3433.,12.0,70,1,"amc rebel sst"
17.0,8,302.0,140.0,3449.,10.5,70,1,"ford torino"
15.0,8,429.0,198.0,4341.,10.0,70,1,"ford galaxie 500"

auto-mpg.xlsx
auto-mpg.csv

.
. .
.

1. Pandas 소개

1.1 Pandas 기초

1.2 Pandas를 활용한 UCI 자동차 연비 데이터 분석

2. Matplotlib 소개

2.1 Matplotlib를 활용한 데이터 분석

2.2 Pandas와 Matplotlib를 활용한 남북한 발전 전력량 데이터 분석

3. Seaborn 소개

4. Numpy 소개

5. Exercise

1. Pandas 소개

1.1 Pandas 기초

1.2 Pandas를 활용한 UCI 자동차 연비 데이터 분석

2. Matplotlib 소개

2.1 Matplotlib를 활용한 데이터 분석

2.2 Pandas와 Matplotlib를 활용한 남북한 발전 전력량 데이터 분석

3. Seaborn 소개

4. Numpy 소개

5. Exercise

■ Pandas 기반의 통계 처리 실습

- 12_(6 page) 강의용 Pandas를 활용한 UCI 자동차 연비 데이터 분석.ipynb

<https://matplotlib.org/>

1. Pandas 소개

1.1 Pandas 기초

1.2 Pandas를 활용한 UCI 자동차 연비 데이터 분석

2. Matplotlib 소개

2.1 Matplotlib를 활용한 데이터 분석

2.2 Pandas와 Matplotlib를 활용한 남북한 발전 전력량 데이터 분석

3. Seaborn 소개

4. Numpy 소개

5. Exercise

■ Matplotlib

- Python에서 데이터를 차트나 플롯으로 시각화하는 라이브러리
- 정형화된 차트나 플롯 이외에도 저수준 api를 사용한 다양한 시각화 기능을 제공

Use Guide

https://matplotlib.org/stable/gallery/lines_bars_and_markers/categorical_variables.html

https://matplotlib.org/stable/gallery/images_contours_and_fields/image_annotated_heatmap.html

1. Pandas 소개

1.1 Pandas 기초

1.2 Pandas를 활용한 UCI 자동차 연비 데이터 분석

2. Matplotlib 소개

2.1 Matplotlib를 활용한 데이터 분석

2.2 Pandas와 Matplotlib를 활용한 남북한 발전 전력량 데이터 분석

3. Seaborn 소개

4. Numpy 소개

5. Exercise

- Matplotlib.pyplot 기반의 데이터 분석 결과 시각화 실습
 - 12_(8 page) 강의용 Matplotlib를 활용한 데이터 분석.ipynb

<http://kosis.kr/index/index.do>

1. Pandas 소개

1.1 Pandas 기초

1.2 Pandas를 활용한 UCI 자동차 연비 데이터 분석

2. Matplotlib 소개

2.1 Matplotlib를 활용한 데이터 분석

2.2 Pandas와 Matplotlib를 활용한 남북한 발전 전력량 데이터 분석

3. Seaborn 소개

4. Numpy 소개

5. Exercise

통계청에서 제공하는 남북한 발전 전력량 데이터셋

	A	B	C	D	E	F	G	H	I
1	전력량 (억kWh)	발전 전력별	1990	1991	1992	1993	1994	1995	1996
2	남한	합계	1,077	1,186	1,310	1,444	1,650	1,847	2,055
3		수력	64	51	49	60	41	55	52
4		화력	484	573	696	803	1,022	1,122	1,264
5		원자력	529	563	565	581	587	670	739
6		신재생	-	-	-	-	-	-	-
7	북한	합계	277	263	247	221	231	230	213
8		수력	156	150	142	133	138	142	125
9		화력	121	113	105	88	93	88	88
10		원자력	-	-	-	-	-	-	-

X	Y	Z	AA	AB	AC
2011	2012	2013	2014	2015	2016
4,969	5,096	5,171	5,220	5,281	5,404
78	77	84	78	58	66
3,343	3,430	3,581	3,427	3,402	3,523
1,547	1,503	1,388	1,564	1,648	1,620
-	86	118	151	173	195
211	215	221	216	190	239
132	135	139	130	100	128
79	80	82	86	90	111
-	-	-	-	-	-

남북한발전전력량.xlsx

1. Pandas 소개

1.1 Pandas 기초

1.2 Pandas를 활용한 UCI 자동차 연비 데이터 분석

2. Matplotlib 소개

2.1 Matplotlib를 활용한 데이터 분석

2.2 Pandas와 Matplotlib를 활용한 남북한 발전 전력량 데이터 분석

3. Seaborn 소개

4. Numpy 소개

5. Exercise

■ Pandas와 Matplotlib 기반의 데이터 분석 및 시각화 실습

- 12_(10 page) 강의용 Pandas와 Matplotlib를 활용한 남북한발전전력량 데이터 분석.ipynb

<https://seaborn.pydata.org/>

1. Pandas 소개

1.1 Pandas 기초

1.2 Pandas를 활용한 UCI 자동차 연비 데이터 분석

2. Matplotlib 소개

2.1 Matplotlib를 활용한 데이터 분석

2.2 Pandas와 Matplotlib를 활용한 남북한 발전 전력량 데이터 분석

3. Seaborn 소개

4. Numpy 소개

5. Exercise

■ Seaborn

- Matplotlib에 기반한 Python 데이터 시각화 라이브러리
- 매력적이고 유익한 통계 그래픽을 표현하기 위한 고급 인터페이스를 제공

Use Guide

<https://seaborn.pydata.org/examples/index.html>

<http://numpy.org/>

1. Pandas 소개

1.1 Pandas 기초

1.2 Pandas를 활용한 UCI 자동차 연비 데이터 분석

2. Matplotlib 소개

2.1 Matplotlib를 활용한 데이터 분석

2.2 Pandas와 Matplotlib를 활용한 남북한 발전 전력량 데이터 분석

3. Seaborn 소개

4. Numpy 소개

5. Exercise

■ Numpy

- 행렬 또는 대규모의 **다차원 배열**을 쉽게 처리 할 수 있도록 지원하는 Python 라이브러리
- 데이터 구조 외에도 수치 계산을 위해 효율적인 기능을 제공

Use Guide

<https://numpy.org/devdocs/user/quickstart.html>

1. Pandas 소개

1.1 Pandas 기초

1.2 Pandas를
활용한 UCI
자동차 연비
데이터 분석

2. Matplotlib 소개

2.1 Matplotlib를
활용한 데이터
분석2.2 Pandas와
Matplotlib를
활용한 남북한
발전 전력량
데이터 분석

3. Seaborn 소개

4. Numpy 소개

5. Exercise

- UCI 자동차 연비 데이터셋의 차종과 연비 항목을 이용하여 산점도(Scatter plot) 그래프를 출력하시오.

auto-mpg.xlsx
auto-mpg.csv

