통계분석

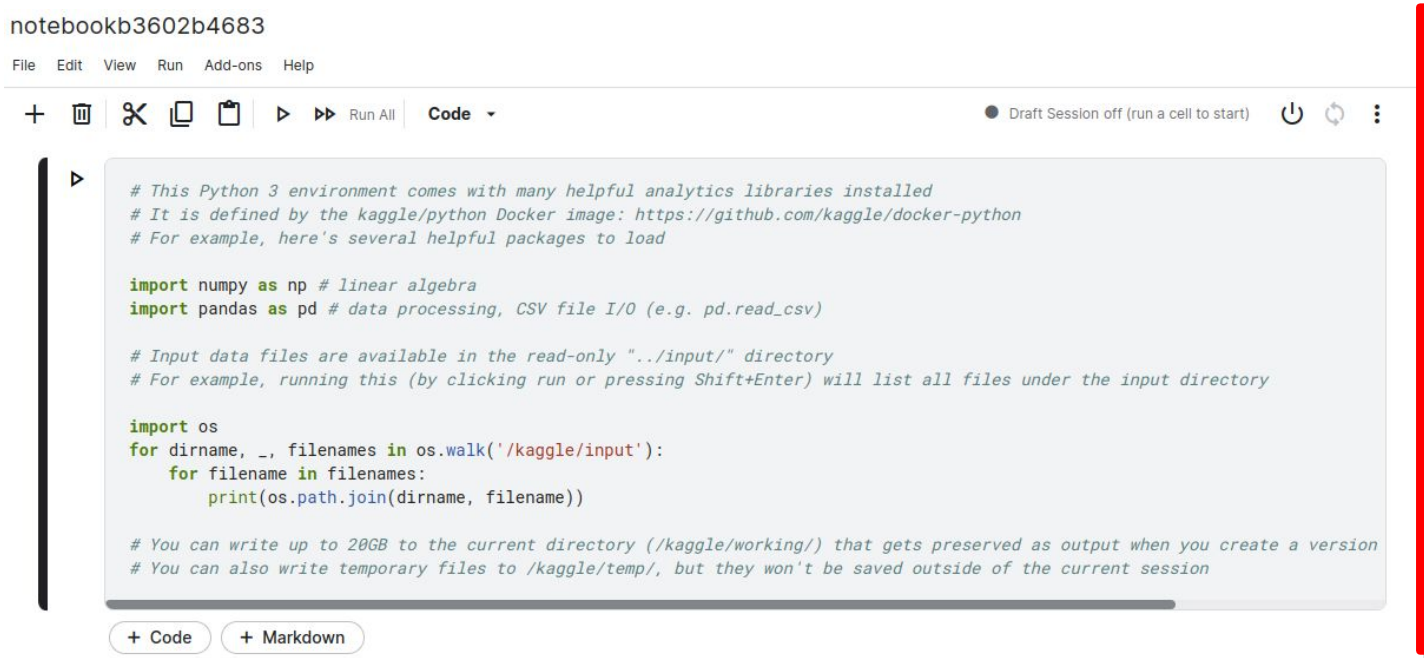
Statistical Analysis

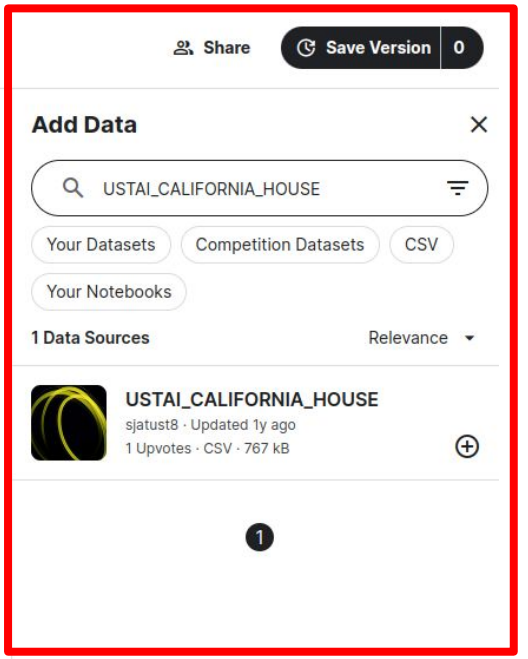
Assignment 01.
Descriptive Statistics

Dataset : California House Prices

In the lecture, we have used a dataset of “Titanic: Machine Learning from Disaster.” in order to practice visualization (matplotlib) and statistical modules (scipy.stats). In this assignment, we will use a different dataset to practice visualization (matplotlib) and statistical modules (scipy.stats). The dataset is about “California House Prices.”

You can add this dataset to a kaggle notebook by searching “USTAI_CALIFORNIA_HOUSE” or by using the URL https://www.kaggle.com/datasets/sjatust8/ustai-california-house.





Dataset : California House Prices

You can read a file “california_house.csv” by using the Pandas module as shown below.

STAT-Assignment01 Draft saved

File Edit View Run Add-ons Help

Share | Save Version 0

Run All | Code | Draft Session (3m)

[2]:
```
import pandas as pd
```

[3]:
```
data = pd.read_csv('/kaggle/input/ustai-california-house/california_house.csv')
```

[4]:
```
data
```

[4]:

	Unnamed: 0	MedInc	HouseAge	AveRooms	AveBedrms	Population	AveOccup	Latitude	Longitude	MedHouseVal
0	0	8.3252	41.0	6.984127	1.023810	322.0	2.555556	37.88	-122.23	4.526
1	1	8.3014	21.0	6.238137	0.971880	2401.0	2.109842	37.86	-122.22	3.585
2	2	7.2574	52.0	8.288136	1.073446	496.0	2.802260	37.85	-122.24	3.521
3	3	5.6431	52.0	5.817352	1.073059	558.0	2.547945	37.85	-122.25	3.413
4	4	3.8462	52.0	6.281853	1.081081	565.0	2.181467	37.85	-122.25	3.422
...	...	...	...	...	...	...	...	...	...	...
20635	20635	1.5603	25.0	5.045455	1.133333	845.0	2.560606	39.48	-121.09	0.781
20636	20636	2.5568	18.0	6.114035	1.315789	356.0	3.122807	39.49	-121.21	0.771
20637	20637	1.7000	17.0	5.205543	1.120092	1007.0	2.325635	39.43	-121.22	0.923
20638	20638	1.8672	18.0	5.329513	1.171920	741.0	2.123209	39.43	-121.32	0.847
20639	20639	2.3886	16.0	5.254717	1.162264	1387.0	2.616981	39.37	-121.24	0.894

20640 rows × 10 columns

Notebook

Data

Add Data

Input

ustai-california-house

california_house.csv

Output (60KB / 19.5GB)

/kaggle/working

Models

Notebook options

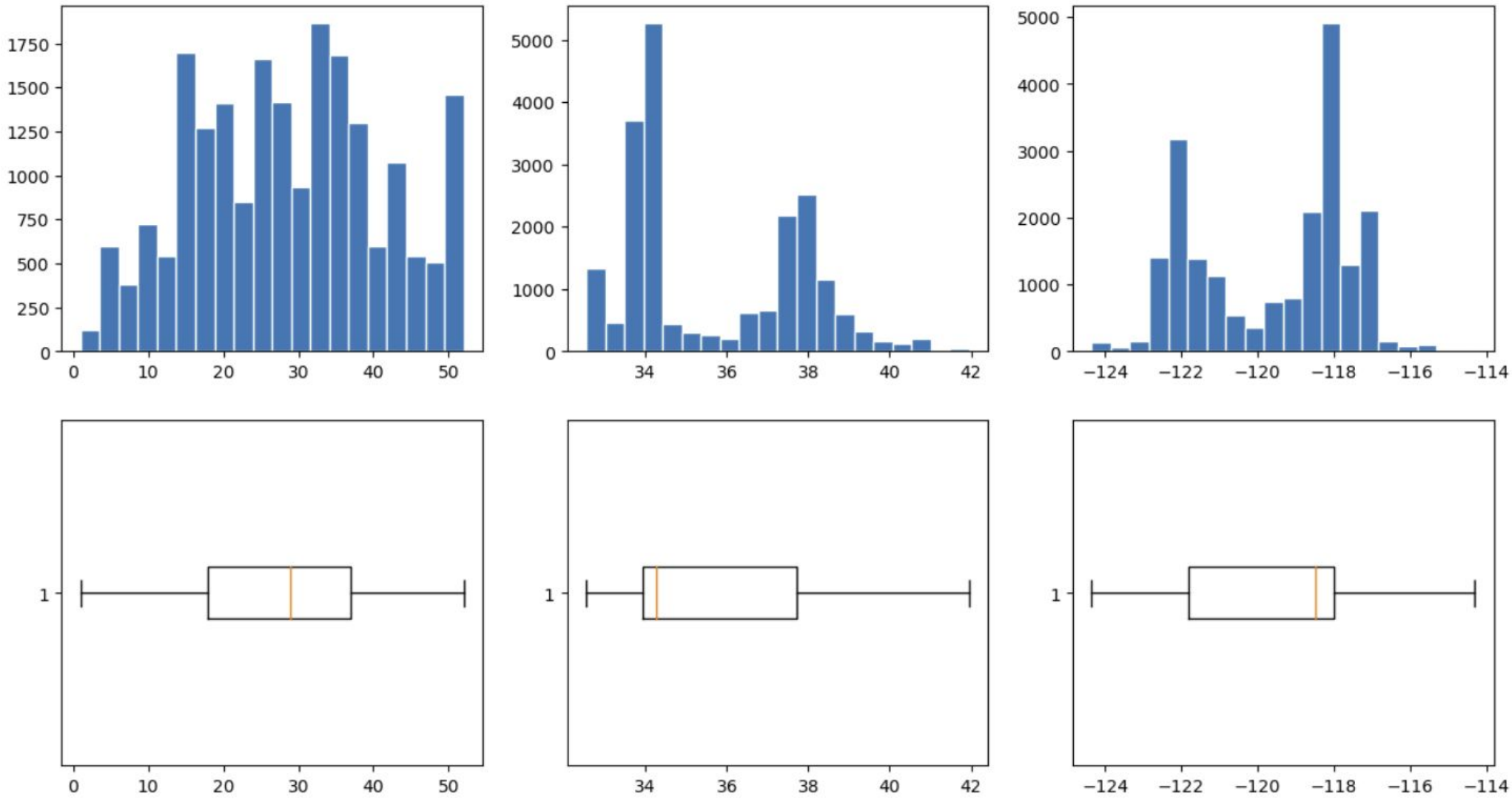
Schedule a notebook to run

Code Help

Find code help

Q1. Drawing Multiple Subplots

For three columns, “HouseAge”, “Latitude”, and “Longitude”, plot the following multiple subplots consisting of histograms and boxplots.



Q2. Calculating Skewness and Kurtosis

For columns from “MedInc” to “MedHouseVal”, calculate skewness and kurtosis. Which column is the most asymmetric? Which column is the sharpest?

Submissions

- Please write down your kaggle notebook file to answer Q1 and Q2. Make a Python code to answer Q1 and Q2, and submit your kaggle notebook file to https://class.ust.ac.kr.
- Due date is 2023. 04. 12.

4Week [22 March - 28 March]

STAT-LEC03-PROBABILITY01

STAT-LEC03-PROBABILITY02

Assignment 01

Assignment 01

Please read the attached file and submit your notebook to solve Q1 and Q2.

STAT_ASSIGNMENT01 20230329.pdf

Submission status

Submission status	No attempt
Grading status	Not graded
Due date	2023-04-12 23:55
Time remaining	14 days 15 hours
Last modified	-
Submission comments	Comments (0)

Add submission