Black Friday - (Assignment 3 - Question 3)

subject Machine Learning / AI

DESCRIPTION

Black Friday falls on the Friday following the 'Thanksgiving Day' and is used as an occasion by many stores to offer highly promoted Sales.

You have the Black Friday dataset, which is an input data file **blackfriday.csv** present at the location **/data/training/blackfriday.csv**

This dataset contains information about purchases made in a retail store on Black Friday sale. Here's a brief description of the columns in the sample dataset:

- **USER_ID**: ID of the user
- **Gender**: F or M
- **Age**: Age group to which the customer belongs
- **Occupation**: ID of occupation of the customer
- **City_Category**: A or B or C
- **Stay_In_Current_City_Years**: 0 to 4+
- **Marital_Status**: 0: Unmarried, 1: Married
- **Purchase**: Purchase amount in dollars

This is a preview of the data under consideration:

User_ID	Gender	Age	Occupation	City_Category	Stay_In_Current_City_Years	Marital_Status	Purchase
1000001	F	0-17	10	A	2	0	8370
1000001	F	0-17	10	A	2	0	15200
1000001	F	0-17	10	A	2	0	1422
1000001	F	0-17	10	A	2	0	1057
1000003	M	26-35	15	A	3	0	15227
1000004	M	46-50	7	B	2	1	19215
1000004	M	46-50	7	B	2	1	15854
1000004	M	46-50	7	B	2	1	15686

The retailer wants to analyse this data and improve its future sales based on the analysis. In all the questions of this Assignment, we have to perform analysis on this data.

Question

- We want to analyse the different statistics of data in **Purchase** column. Obtain the following Summary Statistics for the **Purchase** column and ***print*** their values-

  (**Hint**: Use pandas dataframe functions)

- **Variance**
- **Standard Deviation**
- **Skewness**
- **Kurtosis**

Input Format

- You have to read data from a file named **blackfriday.csv** present at the location **/data/training/blackfriday.csv**

Output Format

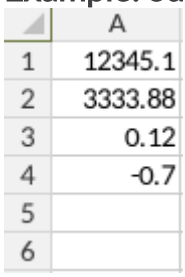
- You have to perform the operations as required by the above question and write (written above as **print**) your output to a file named **output.csv** which should be present at the location **/code/output/output.csv**
- Write only the calculated value of **Variance rounded to 2 decimal places** in the first row
- Write only the calculated value of **Standard Deviation rounded to 2 decimal places** in the second row
- Write only the calculated value of **Skewness rounded to 2 decimal places** in the third row
- Write only the calculated value of **Kurtosis rounded to 1 decimal place** in the fourth row
- Do not write any headers or additional labels in the **output3.csv** file

Sample Input

Read the input file **/data/training/blackfriday.csv**

Sample Output

Example: output.csv will have data looking like this:



	A
1	12345.1
2	3333.88
3	0.12
4	-0.7
5	
6	

DATASETS

- Training dataset

EXECUTION TIME LIMIT

Default.