

School of Computer Science Engineering and Technology

Course- BTech

Course Code- CSET211

Year- Second

Date- 09/08/2022

Type- AI Core-1

Course Name- Statistical Machine Learning

Semester- ODD

Batch- CSE 3rd Semester

Lab Assignment (22nd Aug to 26th Aug 2022)

Lab Assignment- 2 (Topic-Pandas)

Objective: To use Pandas python library and perform various operations on dataset

Diamonds dataset:

This classic dataset contains the prices and other attributes of almost 54,000 diamonds. It's a great dataset for beginners learning to work with data analysis and visualization.



Content

Column Name	Description
price	price in US dollars (\\$326--\\$18,823)
carat	weight of the diamond (0.2--5.01)

cut	quality of the cut (Fair, Good, Very Good, Premium, Ideal)
color	diamond colour, from J (worst) to D (best)
clarity	a measurement of how clear the diamond is (I1 (worst), SI2, SI1, VS2, VS1, VVS2, VVS1, IF (best))
x	length in mm (0--10.74)
y	width in mm (0--58.9)
z	depth in mm (0--31.8)
depth	total depth percentage = $z / \text{mean}(x, y) = 2 * z / (x + y)$ (43--79)
table	width of top of diamond relative to widest point (43--95)

Access dimond.csv :

```
import pandas as pd
```

```
diamonds = pd.read_csv('https://raw.githubusercontent.com/mwaskom/seaborn-data/master/diamonds.csv')
```

```
print(diamonds)
```

Question-1: Create a programme that reads a csv file from a specified source and prints the first 5 rows using the pandas library.

Question 2: Create a program that reads a dataset from the diamonds DataFrame, modifies the default column values, and prints the first six rows.

Question-3: Write a program to select a series from diamonds DataFrame. Print the content of the series.

Question-4: Create a new 'Quality-color' Series of diamond DataFrame by writing a python program.

Question-5: Create a program to determine the number of rows and columns, as well as the data type, of each column in the diamonds Dataframe.

Question-6: Create a program that summarises only the diamonds Dataframe's 'object' columns.

Question-7: Create a Pandas program to read rows in which the 'cut' is 'Premium', column 'color' of diamonds DataFrame.

Question-8: Concatenate the diamonds DataFrame with the 'colour' Series using a Python program.

Question-9. Create a program to calculate the multiply of length, width and depth for each cut of diamonds DataFrame.

Question-10: Create a program to check the number of rows and columns and drop those row if 'any' values are missing in a row of diamonds DataFrame.

Question-11: Creat a program to drop a row if any or all values in a row are missing of diamonds DataFrame on two specific columns.