## Machine Learning Laboratory

## (410302)

## BE Sem I Honors in AI/ML

### Academic Year: 2023-24

Lab Assignment No.3

**Problem Statement:**

Write a program to solve assignments on pandas basic

**Objective**

## **Pandas Lab Exercise -1**

**Diamonds:**

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 / 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)

Copy

Source: https://www.kaggle.com/shivam2503/diamonds

**1.** Write a Pandas program to read a csv file from a specified source and print the first 5 rows.

**2.** Write a Pandas program to read a dataset from diamonds Data Frame and modify the default columns values and print the first 6 rows.

**3.** Write a Pandas program to select a series from diamonds Data Frame. Print the content of the series. 

**4.** Write a Pandas program to create a new 'Quality -color' Series (use bracket notation to define the Series name) of the diamonds Data Frame.

**5.** Write a Pandas program to find the number of rows and columns and data type of each column of diamonds Data Frame.

**6.** Write a Pandas program to summarize only 'object' columns of the diamonds Data Frame. 

**7.** Write a Pandas program to rename two of the columns of the diamonds Data Frame. 

**8.** Write a Pandas program to rename all the columns of the diamonds Data Frame. 

**9.** Write a Pandas program to remove the second column of the diamonds Data Frame. 

**10.** Write a Pandas program to remove multiple columns at once of the diamonds Data Frame. 

**11.** Write a Pandas program to remove multiple rows at once (axis=0 refers to rows) from diamonds Data Frame. 

**12.** Write a Pandas program to sort the 'cut' Series in ascending order (returns a Series) of diamonds Data Frame. 

**13.** Write a Pandas program to sort the 'price' Series in descending order (returns a Series) of diamonds Data Frame. 

**14.** Write a Pandas program to sort the entire diamonds Data Frame by the 'carat' Series in ascending and descending order. 

**15.** Write a Pandas program to filter the Data Frame rows to only show carat weight at least 0.3. 

**16.** Write a Pandas program to convert a python list to pandas’ series. 

**17.** Write a Pandas program to find the details of the diamonds where length>5, width>5 and depth>5. 

**18.** Write a Pandas program to find the diamonds that are either Premium or Ideal. 

**19.** Write a Pandas program to find the diamonds that are with a Fair or Good or Premium. 

**20.** Write a Pandas program to display all column labels of diamonds Data Frame. 

**21.** Write a Pandas program to read only a subset of 3 rows from diamonds Data Frame. 

**22.** Write a Pandas program to iterate through diamonds Data Frame. 

**23.** Write a Pandas program to drop all non-numeric columns from diamonds Data Frame. 

**24.** Write a Pandas program to include only numeric columns in the diamonds Data Frame. 

**25.** Write a Pandas program to pass a list of data types to only describe certain types of diamonds Data Frame. 

**26.** Write a Pandas program to calculate the mean of each numeric column of diamonds Data Frame. 

**27.** Write a Pandas program to calculate the mean of each row of diamonds Data Frame. 

**28.** Write a Pandas program to calculate the mean of price for each cut of diamonds Data Frame. 

**29.** Write a Pandas program to calculate count, minimum, maximum price for each cut of diamonds Data Frame. 

**30.** Write a Pandas program to create a side-by-side bar plot of the diamonds Data Frame. 

**31.** Write a Pandas program to count how many times each value in cut series of diamonds Data Frame occurs. 

**32.** Write a Pandas program to display percentages of each value of cut series occurs in diamonds Data Frame. 

**33.** Write a Pandas program to display the unique values in cut series of diamonds Data Frame. 

**34.** Write a Pandas program to count the number of unique values in cut series of diamonds Data Frame. 

**35.** Write a Pandas program to compute a cross-tabulation of two Series in diamonds Data Frame. 

**36.** Write a Pandas program to calculate various summary statistics of cut series of diamonds Data Frame. 

**37.** Write a Pandas program to create a histogram of the 'carat' Series (distribution of a numerical variable) of diamonds Data Frame. 

**38.** Write a Pandas program to create a bar plot of the 'value\_counts' for the 'cut' series of diamonds Data Frame. 

**39.** Write a Pandas program to create a Data Frame of Booleans (True if missing, False if not missing) from diamonds Data Frame. 

**40.** Write a Pandas program to count the number of missing values in each Series of diamonds Data Frame. 

**41.** Write a Pandas program to check the number of rows and columns and drop those rows if 'any' values are missing in a row of diamonds Data Frame. 

**42.** Write a Pandas program to drop a row if any or all values in a row are missing of diamonds Data Frame on two specific columns. 

**43.** Write a Pandas program to set an existing column as the index of diamonds Data Frame. 

**44.** Write a Pandas program to set an existing column as the index of diamonds Data Frame and restore the index name, and move the index back to a column. 

**45.** Write a Pandas program to access a specified Series index and the Series values of diamonds Data Frame. 

**46.** Write a Pandas program to sort a Series by its values and index of diamonds Data Frame. 

**47.** Write a Pandas program to calculate the multiply of length, width and depth for each cut of diamonds Data Frame. 

**48.** Write a Pandas program to concatenate the diamonds Data Frame with the 'color' Series. 

**49.** Write a Pandas program to read specified rows and all columns of diamonds Data Frame. 

**50.** Write a Pandas program to read rows 0, 5, 7 and all columns of diamonds Data Frame. 

**51.** Write a Pandas program to read rows 2 through 5 and all columns of diamonds Data Frame. 

**52.** Write a Pandas program to read rows 0 through 2 (inclusive), columns 'color' and 'price' of diamonds Data Frame. 

**53.** Write a Pandas program to read rows 0 through 2 (inclusive), columns 'color' through 'price' (inclusive) of diamonds Data Frame. 

**54.** Write a Pandas program to read rows in which the 'cut' is 'Premium', column 'color' of diamonds Data Frame. 

**55.** Write a Pandas program to read rows in positions 0 and 1, columns in positions 0 and 3 of diamonds Data Frame. 

**56.** Write a Pandas program to read rows in positions 0 through 4, columns in positions 1 through 4 of diamonds Data Frame. 

**57.** Write a Pandas program to read rows in positions 0 through 4 (exclusive) and all columns of diamonds Data Frame. 

**58.** Write a Pandas program to read rows 2 through 5 (inclusive), columns in positions 0 through 2 (exclusive) of diamonds Data Frame. 

**59.** Write a Pandas program to print a concise summary of diamonds Data Frame. 

**60.** Write a Pandas program to get the true memory usage by diamonds Data Frame. 

**61.** Write a Pandas program to calculate the memory usage for each Series (in bytes) of diamonds Data Frame. 

**62.** Write a Pandas program to get randomly sample rows from diamonds Data Frame. 

**63.** Write a Pandas program to get sample 75% of the diamonds Data Frame's rows without replacement and store the remaining 25% of the rows in another Data Frame. 

**64.** Write a Pandas program to read the diamonds Data Frame and detect duplicate color.   
Note: duplicated () function returns boolean Series denoting duplicate rows, optionally only considering certain columns.

**65.** Write a Pandas program to count the duplicate rows of diamonds Data Frame.