**Question 01.**

To create DataFrame from dict of narray/list, all the narray must be of same length. If index is passed then the length index should be equal to the length of arrays. If no index is passed, then by default, index will be range(n) where n is the array length.

# Python code demonstrate creating

# DataFrame from dict narray / lists

# By default addresses.

import pandas as pd

# intialise data of lists.

data = \_\_\_\_\_\_\_\_\_\_\_\_\_\_

# Create DataFrame

df = \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

# Print the output.

Print \_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Question 02.**

Read Employees.csv from your localhost the data frame is filtered on the basis of Gender as well as Team. Rows having Gender=”Female” and Team=”Engineering”, “Distribution” or “Finance” are returned.

# importing pandas package

# making data frame from csv file

data = \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

# creating filters of bool series from isin()

# displaying data with both filter applied and mandatory

Print (\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_)

**Question 03.**

Read nba.csv file from your local host. Team name is made as the index column and one team name is passed to .loc method to check if all values with same team name have been returned or not.

# importing pandas package

import pandas as pd

# making data frame from csv file

data = \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

# retrieving rows by loc method

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

# checking data type of rows

print(\_\_\_\_\_\_\_\_\_\_\_)

# display

Print(\_\_\_\_\_\_\_\_\_\_)

**Question 04.**

Read employees.csv file. Largest 5 values are extracted and then compared to the other sorted by the sort\_values() function. NaN values are removed before trying this method.

# importing pandas package

# making data frame from csv file

data =

# removing null values

# extracting greatest 5

large5 = \_\_\_\_\_\_\_\_\_\_\_\_\_

# display

print(large5)

**Question 05.**

rows having all values will be removed. Since the csv file isn’t having such a row, a random row is duplicated and inserted in data frame first.

|  |
| --- |
| #importing pandas package    # making data frame from csv file  data = pd.read\_csv("employees.csv")    #length before adding row  length1 = \_\_\_\_\_\_\_\_\_\_\_\_\_\_    # manually inserting duplicate of a row of row 440  data.loc[1001] = [data["First Name"][440],                    data["Gender"][440],                    data["Start Date"][440],                    data["Last Login Time"][440],                    data["Salary"][440],                    data["Bonus %"][440],                    data["Senior Management"][440],                    data["Team"][440]]      # length after adding row  length2=  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_    # sorting by first name  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  # dropping duplicate values  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  # length after removing duplicates  length3=\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_    # printing all data frame lengths  print(length1, length2, length3) |