1.What is pandas?

Pandas is an open-source data analysis and manipulation library for python.

2.How to install pandas module?

Go to cmd prompt if windows OS.

In cmd type: python -m pip install pandas

It will install required library module.

[Pre-requisite: Needs python to be pre-installed on your machine]

3.How to check whether pandas is installed or not after installation?

Once installation is done via required command, Open VSCode. (If vscode is already opened, then close and re-open it again).

Write a snippet of python code as below:

import pandas as pd

print(“Testing”)

Once you run this script, then you should be able to see the output as “Testing” in the console of vscode. Then, you can confirm pandas is installed successfully.

4.How to Check the pandas version installed?

import pandas as pd

print(pd.\_\_version\_\_)

5.What is the other / alternative way to check pandas version?

pip list

This command in cmd prompt will list all the modules installed along with the version installed on the current system

6.How to Check all in-built classes present in pandas?

Using module inspect2

From cmd prompt type: python -m pip install inspect2

Once installed, close and re-open the VSCode.

import pandas as pd

import inspect

classes = inspect.getmembers(pd, inspect.isclass)

for x,y in classes:

print(x)

7.How to convert a list into a dataframe?

Assume:

import pandas as pd

student\_details=[[1,30],[2,35],[3,21],[4,34]]

def create\_dataframe(student\_details):

column\_name=['student\_id','age']

df=pd.DataFrame(student\_details, columns=column\_name)

#OR

#df=pd.DataFrame(student\_details, columns=('student\_id','age'))

print(df)

create\_dataframe(student\_details)

8. How to return count of total number of rows and columns using pandas?

Using the shape[0] returns n.o of rows and shape[1] returns n.o of columns

Here’s a snippet below:

import pandas as pd

student\_details=[[1,30],[2,35],[3,21],[4,34]]

def create\_dataframe(student\_details):

column\_name=['student\_id','age']

df=pd.DataFrame(student\_details, columns=column\_name)

"""

#OR

print(len(df))

print(len(df.columns))

"""

y=[df.shape[0],df.shape[1]]

#print(y)

return y

create\_dataframe(student\_details)

9. How to return the first 4 rows of data using pandas?

Using head(4). Default head() returns 5 rows of data

Here’s a snippet below:

import pandas as pd

student\_details=[[1,30],[2,35],[3,21],[4,34]]

def create\_dataframe(student\_details):

column\_name=['student\_id','age']

df=pd.DataFrame(student\_details, columns=column\_name)

y=df.head(3)

print(y)

return y

create\_dataframe(student\_details)

10. How to add column label and row label to the 2D data in pandas?

columns and index are the attributes through which the column names/labels and row names/labels can be passed to the DataFrame class

Here’s a snippet below:

import pandas as pd

student\_details=[[1,30],[2,35],[3,21],[4,34]]

def create\_dataframe(student\_details):

column\_name=['student\_id','age']

row\_label=['student1','student2','student3','student4']

df=pd.DataFrame(student\_details, columns=column\_name, index=row\_label)

print(df)

create\_dataframe(student\_details)

11. How to get a particular data from the pandas data set?

Using loc. Suppose if we want to fetch age of student whose student\_id=4,

Here’s a snippet below:

import pandas as pd

student\_details=[[1,30],[2,35],[3,21],[4,34]]

def create\_dataframe(student\_details):

column\_name=['student\_id','age']

row\_label=['student1','student2','student3','student4']

df=pd.DataFrame(student\_details, columns=column\_name, index=row\_label)

print(df)

y=df.loc[df['student\_id']==4,['age']]

print(y)

return y

create\_dataframe(student\_details)

12.How to add a column to an existing data in pandas?

Here’s a snippet below:

import pandas as pd

student\_details=[[1,30],[2,35],[3,21],[4,34]]

def create\_dataframe(student\_details):

column\_name=['student\_id','age']

row\_label=['student1','student2','student3','student4']

df=pd.DataFrame(student\_details, columns=column\_name, index=row\_label)

df['double\_age'] = df['age']\*2

print(df)

create\_dataframe(student\_details)

13.How to remove duplicate rows from a pandas data set?

Using the remove\_duplicates().

{There are options ex:

subset= column\_name #Mention the column name from where you want to remove subsets.

keep= ‘first’ #keeps the first occurrence. If nothing is mentioned, this is default set to first.

Keep = ‘last’ #keeps the last occurrence.}

Here’s a snippet below:

import pandas as pd

customers= [[1, 'Bella', 'bella@example.com'] ,

[2, 'Peter', 'peter@example.com'],

[3, 'Mark', 'mark@example.com'] ,

[4, 'Tom', 'tom@example.com'] ,

[5, 'Harry', 'tom@example.com'] ,

[6, 'Angel', 'angel@example.com']]

column\_names='customer\_id','name','email'

df = pd.DataFrame(customers, columns=column\_names)

df=df.drop\_duplicates(subset='email')

print(df)

14. How to remove missing data from pandas data set?

Two ways:

dropna(): removes the rows if axis=0 and if axis=1, then columns.

notna(): filters and keeps the non-missing values in the data frame.

Here’s a snippet below:

import pandas as pd

students =[[ 'student\_id' , 'name' , 'age' ] ,

[ 355, None, 9 ],

[ 951, None, 8 ],

[ 470, 'Quincy', 20 ],

[ 977, 'Sophia', None ],

[ 300, 'Liam', 15]]

def dataframes(students):

df=pd.DataFrame(students[1:],columns= students[0])

print(df.columns)

df\_cleaned = df[df['name'].notna()]

print(df\_cleaned)

return df\_cleaned

dataframes(students)

15. How to modify columns in pandas data frame?

Assume, if we want to multiply a column (or double it)

Here’s a snippet below:

import pandas as pd

employees=[['name', 'salary' ],

[ 'Jack', 19666 ],

[ 'Piper', 74754 ],

[ 'Mia', 62509 ],

[ 'Ulysses', 54866]]

def modifySalaryColumn(employees: pd.DataFrame) -> pd.DataFrame:

df=pd.DataFrame(employees[1:],columns=employees[0])

df['salary']=df['salary'] \*2

df\_final = df[['name','salary']]

#print(df\_new)

print(df\_final)

#return df\_new

modifySalaryColumn(employees)

16. How to rename column name?

Using 2 ways: Assume employee has column names as [[‘id’],[‘name’],[‘age’]]

First way:

df=pd.DataFrame(employee)

df=df.rename(columns={‘id’:’new\_id’,’name’:’new\_name’,’age’:’new\_age’})

Second way(directly):

df.columns=[‘new\_id’,’new\_name’,’new\_age’]

17. How to rename row name?

Using 2 ways: Assume employee has row names as [[‘id’],[‘name’],[‘age’]]

First way:

df=pd.DataFrame(employee)

df=df.rename(index={‘id’:’new\_id’,’name’:’new\_name’,’age’:’new\_age’})

Second way(directly):

df.index=[‘new\_id’,’new\_name’,’new\_age’]

18. What are the data types in pandas?

There are a total of 8 data types in pandas.

int : 32 and 64 bit integer numbers without decimals.

float: 32 and 64 bit floating numbers with decimals

bool: True and False. It is used for binary values or logical values.

object (a.k.a String): String or text. (Mixed data type)

datetime: datetime64[ns]. Represents data and time values. Suitable for

timedate: timedelta64[ns]. Representing duration or differences between date time values.

categorical data: includes categories such as [‘small’, ‘medium’, ‘large’].

complex data: Involves complex operation 64 and 128 bits.

19. How to change data type in pandas?

Using astype(dtype) for int float and str.

Where, dtype can be int, float or str.

Inside astype(int) converts to int format. Whatever data type is mentioned in astype(), subsequently it gets converted into.

Here’s a snippet below:

import pandas as pd

students=[['student\_id','name','age','grade'],

[1,'Ava',6,73.0],

[2,'Kate',15,87.0]]

def create\_df(students):

df = pd.DataFrame(students[1:], columns=students[0])

df['grade'] = df['grade'].astype(int)

print(df)

create\_df(students)

20. How to change data type for datetime, timedelta and to numeric types in pandas?

Using pd.to\_datetime(), pd.to\_timedelta() and pd.to\_numeric() respectively.

df['numeric\_column'] = pd.to\_numeric(df['numeric\_column'], errors='coerce')

# Convert non-numeric to NaN

df['numeric\_column'] = pd.to\_numeric(df['numeric\_column'], errors='ignore')

# Ignore errors

21. How to change the data type for rows and columns directly?

df.index = df.index.astype(str)

df.columns = df.columns.astype(str)

22.How to know the data type of the pandas data set?

print(df.dtypes)

23. How to fill missing data from None to Zero

Using fillna(number)

fillna() is in-built method.

Number can be any required number.

Here’s a snippet below:

products=[['name' ,'quantity','price'],

[ 'Wristwatch' , None , 135 ],

[ 'WirelessEarbuds' , None , 821 ],

[ 'GolfClubs' , 779 , 9319 ] ,

[ 'Printer' , 849 , 3051 ]]

import pandas as pd

def fillMissingValues(products: pd.DataFrame) -> pd.DataFrame:

df=pd.DataFrame(products[1:],columns=products[0])

df=df.fillna(0)

print(df)

fillMissingValues(products)

24.Can we use fillna() to fill all columns or specific columns?

By default fillna() will