***Pandas Assignment***

**Q1. How do you load a CSV file into a Pandas DataFrame?**

To load a CSV file into the pandas data frame first import the panda library then we have to read the CSV file using pd.read\_csv(' filename.csv ')

Eg. For suppose we have to import or load a CSV file how can we actually do that let us see from the example

**import pandas as pd**

**df = pd.read\_csv(' filen\_name.scv')**

**df**

**Q2. How do you check the data type of a column in a Pandas DataFrame?**

We can check the data type of a particular column and all the columns using .dtypes method

1. If we have to check the data type of an individual column then we have to write df.dtypes['Column\_name']
2. If we have to check data types of entire data frame then we have to simply write df.dtypes

**import pandas as pd**

**employees = [('Stuti', 28, 'Varanasi', 20000),**

**('Saumya', 32, 'Delhi', 25000),**

**('Aaditya', 25, 'Mumbai', 40000),**

**('Saumya', 32, 'Delhi', 35000)**

**]**

**df = pd.DataFrame(employees,**

**columns = ['Name', 'Age', 'City', 'Salary'])**

**datatypes = df.dtypes['Salary']**

**datatypes**

From this we will get the data type of column named Salary and if we want to get the datatypes of entire data frame then we have to simply write df.dtypes

**Q3. How do you select rows from a Pandas DataFrame based on a condition?**

**import pandas as pd**

**employees = [('Stuti', 28, 'Varanasi', 20000),**

**('Saumya', 32, 'Delhi', 25000),**

**('Aaditya', 25, 'Mumbai', 40000),**

**('Saumya', 32, 'Delhi', 35000)**

**]**

**df = pd.DataFrame(employees,**

**columns = ['Name', 'Age', 'City', 'Salary'])**

# selecting rows based on condition

**new\_df= df[df['Salary']>20000]**

**new\_df**

**Q4. How do you rename columns in a Pandas DataFrame?**

We can change column name using .rename method let us understand by an example :

**import pandas as pd**

**employees = [('Stuti', 28, 'Varanasi', 20000),**

**('Saumya', 32, 'Delhi', 25000),**

**('Aaditya', 25, 'Mumbai', 40000),**

**('Saumya', 32, 'Delhi', 35000)**

**]**

**df = pd.DataFrame(employees,**

**columns = ['Name', 'aGe', 'City', 'Salary'])**

*# selecting rows based on condition*

**df.rename(columns = {'aGe':'Age'}, inplace = True)**

**df**

**Q5. How do you drop columns in a Pandas DataFrame?**

We can drop the column by using .drop method

df.drop('Column\_name',axis=1,inplace=True)

Here 'column\_name' is the column that we wish to be dropped from the dataframe.

**Q6. How do you find the unique values in a column of a Pandas DataFrame?**

We can print unique values in the datframe using .unique() method

Also .nunique() is use to calculate number of unique values

**import pandas as pd**

**data = {**

**'A':['21', '2', '3', '24', '5'],**

**'B':['1', '2', '23', '46', '4'],**

**'C':['61', '22', '3', '3', '73'],**

**'D':['1', '23', '2', '52', '2'],**

**'E':['12', '1', '51', '61', '1'] }**

**df = pd.DataFrame(data)**

**#to print number of unique values**

**print(df.A.nunique())**

**#to print number of unique values**

**print(df.A.unique())**

**Output :- 5**

**['21' '2' '3' '24' '5']**

**Q7. How do you find the number of missing values in each column of a Pandas DataFrame?**

We can find number of missing values inside the dataframe using

df.isnull().sum()

**Q8. How do you fill missing values in a Pandas DataFrame with a specific value?**

We can fill nan values in the dataframe using fillna() method

**import pandas as pd**

**import numpy as np**

**data = {**

**'A':['21', '2', np.nan, '24', '5'],**

**'B':['1', '2', '23', '46', '4'],**

**'C':['61', '22', '3', '3', '73'],**

**'D':['1', '23', '2', '52', '2'],**

**'E':['12', '1', '51', '61', '1'] }**

**df = pd.DataFrame(data)**

**n\_df = df.fillna(0)**

**n\_df**

**Q9. How do you concatenate two Pandas DataFrames?**

We can concatenate two dataframe in pandas using pd.cooncat() method

**import pandas as pd**

**df1 = pd.DataFrame({'id': ['1', '2', '3', '4'],**

**'Name': ['ABC', 'TUV', 'MNO', 'GHI']})**

**df2 = pd.DataFrame({'id': ['5', '6', '7', '8'],**

**'Name': ['XYZ', 'PQR', 'DEF', 'JKL']})**

**display(pd.concat([df1, df2]))**

**Q10. How do you merge two Pandas' DataFrames on a specific column?**

**import pandas as pd**

**df1 = pd.DataFrame({'Name':['Unmesh', 'Ashu', 'Ram', 'Sita', 'Gaurav'],**

**'Marks':[90, 87, 75, 88, 89]})**

**df2 = pd.DataFrame({'Name':['Unmesh', 'Divya', 'Ram', 'Gaurav'],**

**'Grade':['A+', 'B++', 'B', 'A'],**

**'Rank':[1, 3, 4, 2 ],**

**'Gender':['Male', 'Female', 'male', 'male']})**

**display(df1)**

**display(df2)**

**df1.merge(df2[['Grade', 'Name']], on = 'Name', how = 'left')**

In the resultant dataframe Grade column of df2 is merged with df1 based on key column Name with merge type left i.e. all the values of left dataframe (df1) will be displayed.

**Q11. How do you group data in a Pandas DataFrame by a specific column and apply an aggregation function?**

**I was able to solve this problem partially.**

**Q12. How do you pivot a Pandas DataFrame?**

pandas.pivot(index, columns, values) function produces pivot table based on 3 columns of the DataFrame. Uses unique values from index / columns and fills with values.

**Q13. How do you change the data type of a column in a Pandas DataFrame?**

astype() function is used to change the datatype of a column into another one

**Q14. How do you sort a Pandas DataFrame by a specific column?**

To sort the DataFrame based on the values in a single column, you’ll use .sort\_values(). By default, this will return a new DataFrame sorted in ascending order. It does not modify the original DataFrame.

**Q15. How do you create a copy of a Pandas DataFrame?**

The copy() method returns a copy of the DataFrame. By default, the copy is a "deep copy" meaning that any changes made in the original DataFrame will NOT be reflected in the copy.

**Q16. How do you filter rows of a Pandas DataFrame by multiple conditions?**

Let us understand it by an example

import pandas as pd

dataFrame = pd.DataFrame({'Name': ['RACHEL', 'MONICA', 'PHOEBE','ROSS','CHANDLER','JOEY'],

'Age': [30, 35, 37, 33, 34, 30],

'Salary': [100000, 93000, 88000, 120000, 94000, 95000],

'JOB': ['DESIGNER', 'CHEF', 'MASUS', 'PALENTOLOGY','IT', 'ARTIST']})

# filter dataframe based on multiple condition

display(dataFrame.loc[(dataFrame['Salary']>=100000) & (dataFrame['Age']< 40) & (dataFrame['JOB'].str.startswith('D')),

['Name','JOB']])

**Q17. How do you calculate the mean of a column in a Pandas DataFrame?**

**import pandas as pd**

**df = pd.DataFrame({"A":[12, 4, 5, 44, 1],**

**"B":[5, 2, 54, 3, 2],**

**"C":[20, 16, 7, 3, 8],**

**"D":[14, 3, 17, 2, 6]})**

**df.mean()**

**Output :- A 13.2**

**B 13.2**

**C 10.8**

**D 8.4**

**Q18. How do you calculate the standard deviation of a column in a Pandas DataFrame?**

df.std(axis = 0, skipna = True) using this method we can calculate the standard deviation of the numeric column.

Skipna = True means it will automatically skip the nan values and then calculate the standard deviation.

**Q19. How do you calculate the correlation between two columns in a Pandas DataFrame?**

By using corr() function we can get the correlation between two columns in the dataframe.

Correlation is used to summarize the strength and direction of the linear association between two quantitative variables. It is denoted by r and values between -1 and +1.

If the correlation between two columns is near to +1 the stronger the correlation between the two variables the farther the number from +1 the correlation will be weak.

**Q20. How do you select specific columns in a DataFrame using their labels?**

By using df.Column\_name or using df['Column\_name']

Here column\_name means the column which you want to print

**import pandas as pd**

**df = pd.DataFrame({"A":[12, 4, 5, 44, 1],**

**"B":[5, 2, 54, 3, 2],**

**"C":[20, 16, 7, 3, 8],**

**"D":[14, 3, 17, 2, 6]})**

**df['A']**

**Q21. How do you select specific rows in a DataFrame using their indexes?**

Using .iloc method used for selection based on position. It is similar to loc[] indexer but it takes only integer values to make selections.

**import pandas as pd**

**employees = [**

**('Saumya', 32, 'Delhi', 25000),**

**('Aaditya', 25, 'Mumbai', 40000),**

**('Aaditya', 40, 'Dehradun', 24000),**

**('Seema', 32, 'Delhi', 70000)**

**]**

**df = pd.DataFrame(employees,**

**columns =['Name', 'Age',**

**'City', 'Salary'])**

**# Set 'Name' column as index**

**# on a Dataframe**

**df.set\_index("Name", inplace = True)**

**result = df.iloc["2"]**

**result**

**Output :- Age 40**

**City Dehradun**

**Salary 24000**

**Name: Aaditya, dtype: object**

**Q22. How do you sort a DataFrame by a specific column?**

We can sort the dataFrame by specified column by using df.sort\_values(by = 'Column\_Name') here column name is the name of the column which you want to sort

**import pandas as pd**

**employees = [('Sanket',25,'Pune',55000),**

**('Saumya', 32, 'Delhi', 25000),**

**('Aaditya', 25, 'Mumbai', 40000),**

**('Aaditya', 40, 'Dehradun', 24000),**

**('Ram',25,'Pune',95000),**

**('Seema', 32, 'Delhi', 70000)**

**]**

**df = pd.DataFrame(employees,columns =['Name', 'Age','City', 'Salary'])**

**reslt\_df = df.sort\_values(by = 'Name')**

**reslt\_df**

**Q23. How do you create a new column in a DataFrame based on the values of another column?**

**Not able to solve this question**

**Q24. How do you remove duplicates from a DataFrame?**

df.drop\_duplicates(keep=False) method is used to delete the duplicates from a data frame

**Q25. What is the difference between .loc and .iloc in Pandas?**

The .loc[ ] function selects the data by labels of rows or columns. It can select a subset of rows and columns. There are many ways to use this function whereas

The .iloc[ ] is used for selection based on position. It is similar to loc[ ] indexer but it takes only integer values to make selections.