**Pandas Assignment**

Q.1) How do you load a CSV file into a Pandas DataFrame?

Ans.) We have to use read\_csv() function to load the CSV file into a data frame.

Q.2) How do you check the data type of a column in a Pandas DataFrame?

Ans.) To check the data type in pandas DataFrame we can use the “dtype” attribute. The attribute returns a series with the data type of each column. And the column names of the DataFrame are represented as the index of the resultant series object and the corresponding data types are returned as values of the series object.

Q.3) How do you select rows from a Pandas DataFrame based on a condition?

Ans.) There are several ways to select rows from a Pandas dataframe:

1. Boolean indexing ( df[df['col'] == value ] )
2. Positional indexing ( df. iloc[...] )
3. Label indexing ( df. xs(...) )
4. df. query(...) API.

Q.4) How do you rename columns in a Pandas DataFrame?

Ans.) 4 Ways to Rename Pandas Columns

1. Method 1: using rename() function.
2. Method 2: assigning list of new column names.
3. Method 3: replacing the columns string.
4. Method 4: using set\_axis() function.

Q.5) How do you drop columns in a Pandas DataFrame?

Ans.) The drop() method removes the specified row or column. By specifying the column axis ( axis='columns' ), the drop() method removes the specified column. By specifying the row axis ( axis='index' ), the drop() method removes the specified row.

Q.6) How do you find the unique values in a column of a Pandas DataFrame?

Ans.) To reduce your manual work, below are the 5 methods by which you can easily get unique values in a column of pandas dataframe:

1. 1) Using unique() method.
2. 2) Using the drop\_duplicates method.
3. 3) Get unique values in multiple columns.
4. 4) Count unique values in a single column.
5. 5) Count unique values in each columns.

Q.7) How do you find the number of missing values in each column of a Pandas DataFrame?

Ans.) You can use the following syntax to count NaN values in Pandas DataFrame:

1. (1) Count NaN values under a single DataFrame column: df['column name'].isna().sum()
2. (2) Count NaN values under an entire DataFrame: df.isna().sum().sum()
3. (3) Count NaN values across a single DataFrame row: df.loc[[index value]].isna().sum().sum()

Q.8) How do you fill missing values in a Pandas DataFrame with a specific value?

Ans.) The fillna() method replaces the NULL values with a specified value. The fillna() method returns a new DataFrame object unless the inplace parameter is set to True , in that case the fillna() method does the replacing in the original DataFrame instead.

Q.9) How do you concatenate two Pandas DataFrames?

Ans.) We'll pass two dataframes to pd. concat() method in the form of a list and mention in which axis you want to concat, i.e. axis=0 to concat along rows, axis=1 to concat along columns.

Q.10) How do you merge two Pandas DataFrames on a specific column?

1. Ans.) You can join pandas Dataframes in much the same way as you join tables in SQL.
2. The concat() function can be used to concatenate two Dataframes by adding the rows of one to the other.
3. concat() can also combine Dataframes by columns but the merge() function is the preferred way.

Q.11) How do you group data in a Pandas DataFrame by a specific column and apply an aggregation function?

Ans.) In Pandas, we can also apply different aggregation functions across different columns. For that, we need to pass a dictionary with key containing the column names and values containing the list of aggregation functions for any specific column.

Q.12) How do you pivot a Pandas DataFrame?

Ans.) The pivot() function is used to reshaped a given DataFrame organized by given index / column values. This function does not support data aggregation, multiple values will result in a MultiIndex in the columns. Column to use to make new frame's index. If None, uses existing index.

Q.13) How do you change the data type of a column in a Pandas DataFrame?

Ans.) The best way to convert one or more columns of a DataFrame to numeric values is to use pandas.to\_numeric() . This function will try to change non-numeric objects (such as strings) into integers or floating-point numbers as appropriate.

Q.14) How do you sort a Pandas DataFrame by a specific column?

Ans.) 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.

Q.15) How do you create a copy of a Pandas DataFrame?

Ans.) 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.

Q.16) How do you filter rows of a Pandas DataFrame by multiple conditions?

Ans.) The loc function in pandas can be used to access groups of rows or columns by label. Add each condition you want to be included in the filtered result and concatenate them with the & operator. You'll see our code sample will return a pd. dataframe of our filtered rows.

Q.17) How do you calculate the mean of a column in a Pandas DataFrame?

Ans.) To calculate the mean of whole columns in the DataFrame, use pandas. Series. mean() with a list of DataFrame columns. You can also get the mean for all numeric columns using DataFrame.

Q.18) How do you calculate the standard deviation of a column in a Pandas DataFrame?

Ans.) Standard deviation is calculated using the function . std() . However, the Pandas library creates the Dataframe object and then the function . std() is applied on that Dataframe .

Q.19) How do you calculate the correlation between two columns in a Pandas DataFrame?

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

Q.20) How do you select specific columns in a DataFrame using their labels?

Ans.) Selecting columns based on their name

This is the most basic way to select a single column from a dataframe, just put the string name of the column in brackets. Returns a pandas series. Passing a list in the brackets lets you select multiple columns at the same time.

Q.21) How do you select specific rows in a DataFrame using their indexes?

Ans.) In the Pandas DataFrame we can find the specified row value with the using function iloc(). In this function we pass the row number as parameter.

Q.22) How do you sort a DataFrame by a specific column?

Ans.) 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.

Q.23) How do you create a new column in a DataFrame based on the values of another column?

Ans.) You can add/append a new column to the DataFrame based on the values of another column using  df.assign() , df. apply() , and, np. where() functions and return a new Dataframe after adding a new column.

Q.24) How do you remove duplicates from a DataFrame?

Ans.) Pandas DataFrame drop\_duplicates() Method

The drop\_duplicates() method removes duplicate rows. Use the subset parameter if only some specified columns should be considered when looking for duplicates.

Q.25) What is the difference between .loc and .iloc in Pandas?

Ans.) The main distinction between the two methods is: loc gets rows (and/or columns) with particular labels. iloc gets rows (and/or columns) at integer locations.