**Pandas Assignment**

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

A1. To load a CSV file into a Pandas DataFrame, one must use the following code:

import pandas as pd  
  
df = pd.read\_csv('file.csv')

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

A2. 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.

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

A3. To select rows from Dataframe based on conditions, we use different methods such as indexing, loc[], etc. with name of the row inside the brackets and apply '>', '=', '=', '<=', '!=' operators.

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

A4. One can rename the columns in a Pandas Dataframe is by using the rename() function, set\_axis() function or simply by assigning new column names.

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

A5. The drop() method removes the specified row or column. By specifying the column axis (axis='columns'), the drop() method removes the specified column.

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

A6. One can find unique values by unique function in two formats:

1. series. unique() : In this one has to add the unique function after the series(column) in which we want to find the unique values.
2. pd. unique() : In this we have to pass the series as a parameter to find the unique values.

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

A7. One can use the following syntax to find the number of missing values in each column of a Pandas DataFrame : df['column name'].isna().sum()

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

A8. One can use the fillna() function to fill missing values in a Pandas DataFrame with a specific 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.

Q9. How do you concatenate two Pandas DataFrames?

A9. One can use concat() function to concatenate two Pandas DataFrames.

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

A10. To merge two Pandas DataFrames on a specific column, merge() function can be used.

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

aggregation function?

A11. To group data in a Pandas DataFrame by a specific column, one can use DataFrame.groupby() function. This function returns the DataFrameGroupBy object and use aggregate() function to calculate the sum.

Q12. How do you pivot a Pandas DataFrame?

A12. The pivot() function is used to reshaped a given DataFrame organized by given index / column values.

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

A13. To change the data type of a column in a Pandas DataFrame, one can use DataFrame.astype()**.** This function also helps to convert any suitable existing column to a categorical type.

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

A14. One can sort by a Pandas DataFrame by a specific column using sort\_values() method. To specify the order, one has to use ascending boolean property; False for descending and True for ascending. By default, it is set to True.

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

A15. The copy() method returns a copy of the DataFrame.

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

A16. There are few functions to filter rows of a Pandas DataFrame by multiple conditions:

1. loc() function
2. np.where() function
3. query() function
4. eval() function

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

A17. To calculate the mean of a column in the Pandas DataFrame, dataframe.mean() function can be used.

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

A18. To calculate the mean of a column in the Pandas DataFrame, dataframe.std() function can be used.

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

A19. To calculate the correlation between two columns in a Pandas DataFrame, the function corr() is used in the format of name\_of\_dateframe[column1\_name].corr(name\_of\_dateframe[column1\_name]).

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

A20. To select specific columns in a DataFrame using their labels, one can use DataFrame.loc[] and DataFrame.iloc[] to select a single column or multiple columns from pandas DataFrame by column names/label or index position respectively. The loc[] is user defined and the iloc[] is system defined.

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

A21. To select specific rows in a DataFrame using their indexes, one can simply list down the column names.

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

A22. To sort a DataFrame by a specific column, one can use sort\_values() function.

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

A23. To create a new column in a DataFrame based on the values of another column, one can use apply() and map() functions.

Q24. How do you remove duplicates from a DataFrame?

A24. To remove duplicates from a DataFrame, drop\_duplicates() function can be used.

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

A25. The main difference between loc and iloc is that the loc is label-based, which means that you must specify rows and columns based on their row and column labels. iloc is integer position-based, so you must specify rows and columns by their integer position values (0-based integer position).