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

**Ans:** Pandas read\_csv() function imports a CSV file to DataFrame format. header: this allows you to specify which row will be used as column names for your dataframe. Expected an int value or a list of int values. Default value is header=0 , which means the first row of the CSV file will be treated as column names.

**Q2. 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.

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

**Ans:** Let’s see how to Select rows based on some conditions in Pandas DataFrame.

### Selecting rows based on particular column value using '>', '=', '=', '<=', '!=' operator.

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

**Ans: 1)** One way of renaming the columns in a Pandas Dataframe is by using the rename() function.

2) The columns can also be renamed by directly assigning a list containing the new names to the columns attribute of the Dataframe object for which we want to rename the columns.

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

### Ans: 1) We can drop Columns from a Dataframe using drop() method.

### 2) We can Drop Columns from a Dataframe using iloc[] and drop() method.

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

**Ans:** We can find unique values by unique function in two formats:

1. series. unique() : In this we have 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?**

**Ans:** Count missing values in each row and column,sum() calculates the sum of elements for each row and column. Since sum() calculate as True=1 and False=0 , you can count the number of missing values in each row and column by calling sum() from the result of isnull()

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

**Ans:** In Pandas missing data is represented by two value:

* None: None is a Python singleton object that is often used for missing data in Python code.
* NaN : NaN (an acronym for Not a Number), is a special floating-point value recognized by all systems that use the standard IEEE floating-point representation

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

**Ans:** Use pandas.concat() to concatenate/merge two or multiple pandas DataFrames across rows or columns. When you concat() two pandas DataFrames on rows, it creates a new Dataframe containing all rows of two DataFrames basically it does append one DataFrame with another. When you use concat() on columns it performs the join operation.

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

**Ans:** We can merge two Pandas DataFrames on certain columns using the merge function by simply specifying the certain columns for merge.

**Syntax:**DataFrame.merge(right, how=’inner’, on=None, left\_on=None, right\_on=None, left\_index=False, right\_index=False, sort=False, copy=True, indicator=False, validate=None)

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

**Ans:** Dataframe.aggregate() function is used to apply some aggregation across one or more column. Aggregate using callable, string, dict, or list of string/callables. Most frequently used aggregations are:

**sum:** Return the sum of the values for the requested axis  
**min:** Return the minimum of the values for the requested axis  
**max:** Return the maximum of the values for the requested axis

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

**Ans:** 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.

***Parameters:******index[ndarray] :****Labels to use to make new frame’s index****columns[ndarray] :****Labels to use to make new frame’s columns****values[ndarray] :****Values to use for populating new frame’s values*

***Returns:****Reshaped DataFrame****Exception:****ValueError raised if there are any duplicates.*

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

**Ans:** [DataFrame.astype()](https://www.geeksforgeeks.org/python-pandas-dataframe-astype/) method is used to cast pandas object to a specified dtype. This function also provides the capability to convert any suitable existing column to a categorical type.

**Q14. 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.

**Q15. 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.

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

**Ans:** You can select the Rows from Pandas DataFrame based on column values or based on multiple conditions either **using DataFrame.** **loc[] attribute, DataFrame.** **query() or DataFrame.** **apply() method to use lambda function**.

**Q17. 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.mean(), use axis=0 argument to calculate the column-wise mean of the DataFrame.

**Q18. 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 .

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

**Ans:** Initialize two variables, col1 and col2, and assign them the columns that you want to find the correlation of. Find the correlation between col1 and col2 by using df[col1]. corr(df[col2]) and save the correlation value in a variable, corr. Print the correlation value, corr.

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

**Ans:** df.loc[:, 'column1']

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

**Ans:** You can select a single row from pandas DataFrame by integer index using df. iloc[n] . Replace n with a position you wanted to select.

**Q22. 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.

**Q23. 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.

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

**Ans: *Syntax:****DataFrame.drop\_duplicates(subset=None, keep=’first’, inplace=False)*

***Parameters:***

* ***subset:****Subset takes a column or list of column label. It’s default value is none. After passing columns, it will consider them only for duplicates.*
* ***keep:****keep is to control how to consider duplicate value. It has only three distinct value and default is ‘first’.*
  + *If ‘****first****‘, it considers first value as unique and rest of the same values as duplicate.*
  + *If ‘****last****‘, it considers last value as unique and rest of the same values as duplicate.*
  + *If****False****, it consider all of the same values as duplicates*
* ***inplace:****Boolean values, removes rows with duplicates if True.*

***Return type:****DataFrame with removed duplicate rows depending on Arguments passed.*

**Q25. 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.