# PandasAssignment

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

To import a CSV file and put the contents into a Pandas dataframe we use the read\_csv() function, which is appended after calling the pd object we created when we imported Pandas. The read\_csv() function can take several arguments, but by default you just need to provide the path to the file you wish to read

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

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?

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

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

**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.

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

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?

To select the unique values from a specific column in a Pandas dataframe you can use the unique() method. This is simply appended to the end of the column name, e.g. df['column\_name']. unique() and returns a Python list of the unique values.

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

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() . You can count missing values in each column by default, and in each row with axis=1 .

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

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.

Q9. How do you concatenate two Pandas DataFrames?

A simpler way to concatenate multiple DataFrames is to use the concat  function from pandas library.

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

1. merge() for combining data on common columns or indices.
2. .join() for combining data on a key column or an index.
3. concat() for combining DataFrames across rows or columns.

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

To do grouping use DataFrame. groupby() function. This function returns the DataFrameGroupBy object and use aggregate() function to calculate the sum. Similarly, you can also calculate aggregation for all other functions specified in the above table.

Q12. How do you pivot a Pandas DataFrame?

**Create Your Own Pandas Pivot Table in 4 Steps**

1. Download or import the data that you want to use.
2. In the pivot\_table function, specify the DataFrame you are summarizing, along with the names for the indexes, columns and values.
3. Specify the type of calculation you want to use, such as the mean.

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

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.

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

Sorting Your DataFrame on a Single 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.

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

Pandas DataFrame copy() Method  
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?

To filter Pandas DataFrame by multiple conditions use DataFrame. loc[] property along with the conditions. Make sure you surround each condition with a bracket.

.

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

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.

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

First, we'll use the std() method to calculate the standard deviation of all columns in the DataFrame. To do this, you simply append the std() method to the DataFrame object. It returns a Series object with the standard deviation of each column.

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

To calculate the Pearson correlation coefficient for every pair of values in the dataframe, you can simply append the corr() method to the end of the dataframe object. The resulting dataframe, or matrix, will have the correlation coefficient for every pair of columns in the dataframe.

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

**You can use the following methods to select columns by name in a pandas DataFrame:**

1. Method 1: Select One Column by Name df. loc[:, 'column1']
2. Method 2: Select Multiple Columns by Name df. loc[:, ['column1', 'column3', 'column4']]
3. Method 3: Select Columns in Range by Name df. loc[:, 'column2':'column4']

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

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?

Sorting Your DataFrame on a Single 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.

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

We can use **DataFrame.apply()** function to achieve this task.

# importing pandas as pd

import pandas as pd

# Creating the DataFrame

df = pd.DataFrame({'Date':['10/2/2011', '11/2/2011', '12/2/2011', '13/2/2011'],

                    'Event':['Music', 'Poetry', 'Theatre', 'Comedy'],

                    'Cost':[10000, 5000, 15000, 2000]})

# Print the dataframe

print(df)

**Output :**



Now we will create a new column called ‘Discounted\_Price’ after applying a 10% discount on the existing ‘Cost’ column.

* Python3

|  |
| --- |
| # using apply function to create a new column  df['Discounted\_Price'] = df.apply(lambda row: row.Cost -                                    (row.Cost \* 0.1), axis = 1)    # Print the DataFrame after addition  # of new column  print(df) |

**Output :**



**Example 2:**We can achieve the same result by directly performing the required operation on the desired column element-wise.

|  |
| --- |
| import pandas as pd    # Creating the DataFrame  df = pd.DataFrame({'Date':['10/2/2011', '11/2/2011', '12/2/2011', '13/2/2011'],                      'Event':['Music', 'Poetry', 'Theatre', 'Comedy'],                      'Cost':[10000, 5000, 15000, 2000]})    # Create a new column 'Discounted\_Price' after applying  # 10% discount on the existing 'Cost' column.    # create a new column  df['Discounted\_Price'] = df['Cost'] - (0.1 \* df['Cost'])    # Print the DataFrame after  # addition of new column  print(df) |

**Output :**



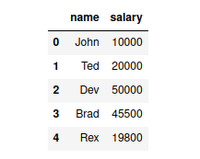
**Example 3:** Using **DataFrame.map()** function to create new column from existing column using a mapping function

We will create a dataframe with some sample data:

* Python3

|  |
| --- |
| data = {      "name": ["John", "Ted", "Dev", "Brad", "Rex", "Smith", "Samuel", "David"],      "salary": [10000, 20000, 50000, 45500, 19800, 95000, 5000, 50000]  }  # create dataframe from data dictionary  df = pd.DataFrame(data)  # print the dataframe  display(df.head()) |

**Output:**



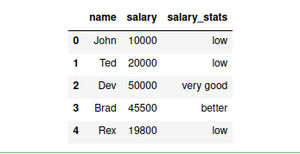
*Sample dataframe*

Now, we will create a mapping function (salary\_stats) and use the DataFrame.map() function to create a new column from an existing column

* Python3

|  |
| --- |
| def salary\_stats(value):      if value < 10000:          return "very low"      if 10000 <= value < 25000:          return "low"      elif 25000 <= value < 40000:          return "average"      elif 40000 <= value < 50000:          return "better"      elif value >= 50000:          return "very good"    df['salary\_stats'] = df['salary'].map(salary\_stats)  display(df.head()) |

**Output:**



**Explanation:** Here we have used pandas DataFrame.map() function to map each value to a string based on our defined mapping logic. The resultant series of values is assigned to a new column, “salary\_stats”.

Q24. How do you remove duplicates from a DataFrame?

You can set 'keep=False' in the drop\_duplicates() function to remove all the duplicate rows. For E.x, df. drop\_duplicates(keep=False)

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

Pandas library of python is very useful for the manipulation of mathematical data and is widely used in the field of machine learning. It comprises many methods for its proper functioning. Loc()and **iloc()** are one of those methods. These are used in slicing data from the [Pandas DataFrame](https://www.geeksforgeeks.org/python-pandas-dataframe/). They help in the convenient selection of data from the DataFrame in [Python](https://www.geeksforgeeks.org/python-programming-language/). They are used in filtering the data according to some conditions.