PANDAS Project Exercise

Guidelines: -

   i.   *This Assignment is mandatory for everyone.*
   ii.   *It is mandatory to submit the answer with the screen shot of the output you have received. Otherwise, no marks will be given.*
   iii.   *If anyone fails to submit the assignment within the last date. His/her assignment submission will not be evaluated and will be allotted minus marks also.*

Questions:

1. How to import pandas and check the version?
2. How to create a series from a list, numpy array and dict?

```
import numpy as np
mylist = list('abcedfghijklmnopqrstuvwxyz') myarr =
np.arange(26)
mydict = dict(zip(mylist, myarr))
```

3. How to convert the index of a series into a column of a dataframe?
4. How to combine many series to form a dataframe?
5. How to assign name to the series' index?
6. How to get the items of series A not present in series B?
7. How to get the items not common to both series A and series B?
8. How to get the minimum, 25th percentile, median, 75th, and max of a numeric series?
9. How to get frequency counts of unique items of a series?
10. How to keep only top 2 most frequent values as it is and replace everything else as 'Other'?
11. How to bin a numeric series to 10 groups of equal size?
12. How to convert a numpy array to a dataframe of given shape? Reshape the series ser into a dataframe with 7 rows and 5 columns

```
ser = pd.Series(np.random.randint(1, 10, 35))
```

13. How to find the positions of numbers that are multiples of 3 from a series?
14. How to extract items at given positions from a series?
    From ser, extract the items at positions in list pos.

```
ser = pd.Series(list('abcdefghijklmnopqrstuvwxyz')) pos = [0, 4, 8,
14, 20]
```

15. How to stack two series vertically and horizontally ?
16. How to convert the first character of each element in a series to uppercase?

```
ser = pd.Series(['how', 'to', 'kick', 'ass?'])
```

17. How to calculate the number of characters in each word in a series?
    Input: ser = pd.Series(['how', 'to', learn, 'pandas?'])

18. How to convert a series of date-strings to a timeseries?
    Input: ser = pd.Series(['01 Jan 2010', '02-02-2011', '20120303', '2013/04/04', '2014-05-05',
    '2015-06-06T12:20'])

    Desired Output:
    0   2010-01-01   00:00:00
    1   2011-02-02   00:00:00

```
2    2012-03-03 00:00:00
3    2013-04-04 00:00:00
4    2014-05-05 00:00:00
5    2015-06-06 12:20:00
dtype: datetime64[ns]
```

19. How to get the day of month, week number, day of year and day of week from a series of date strings?

Input: `ser = pd.Series(['01 Jan 2010', '02-02-2011', '20120303', '2013/04/04', '2014-05-05', '2015-06-06T12:20'])`

Desired Output:
```
Date:  [1, 2, 3, 4, 5, 6]
Week number:  [53, 5, 9, 14, 19, 23]
Day num of year:  [1, 33, 63, 94, 125, 157]
Day of week:  ['Friday', 'Wednesday', 'Saturday', 'Thursday', 'Monday', 'Saturday']
```

20. How to compute the autocorrelations of a numeric series?
21. How to import only specified columns from a csv file?
22. How to rename a specific columns in a dataframe? (Case: Rename the column 'Type' as 'CarType' in df and replace the '.' in column names with '_'.)
23. How to check if a dataframe has any missing values?
24. How to count the number of missing values in each column?
25. How to replace missing values of multiple numeric columns with the mean?(Case: Replace missing values in Min.Price and Max.Price columns of 'df' dataframe with their respective mean.)
26. How to create a primary key index by combining relevant columns? (Case: In df, Replace NaNs with 'missing' in columns 'Manufacturer', 'Model' and 'Type' and create a index as a combination of these three columns and check if the index is a primary key.)
27. How to create one-hot encodings of a categorical variable (dummy variables)? (Case: Get one-hot encodings for column 'a' in the dataframe df and append it as columns.)

Input: `df = pd.DataFrame(np.arange(25).reshape(5,-1), columns=list('abcde'))`

```
     a   b   c   d   e
0    0   1   2   3   4
1    5   6   7   8   9
2   10  11  12  13  14
3   15  16  17  18  19
4   20  21  22  23  24
```

Desired Output:
```
   0  5  10  15  20   b  c  d  e
0  1  0   0   0   0   1  2  3  4
1  0  1   0   0   0   6  7  8  9
```

```
2  0  0  1  0  0  11  12  13  14
3  0  0  0  1  0  16  17  18  19
4  0  0  0  0  1  21  22  23  24
```

28. How to normalize all columns in a dataframe?
29. How to join two dataframes by 2 columns so they have only the common rows?
30. How to remove rows from a dataframe that are present in another dataframe?

```
2  0  0  1  0  0  11  12  13  14
3  0  0  0  1  0  16  17  18  19
4  0  0  0  0  1  21  22  23  24
```