1.

a) Import pandas under the name pd.
b) Print the version of pandas that has been imported.
c) Print out all the version information of the libraries that are required by the pandas library.

2. Consider the following Python dictionary `data` and Python list `labels`:

data = {'animal': ['cat', 'cat', 'snake', 'dog', 'dog', 'cat', 'snake', 'cat', 'dog', 'dog'],

'age': [2.5, 3, 0.5, np.nan, 5, 2, 4.5, np.nan, 7, 3],

'visits': [1, 3, 2, 3, 2, 3, 1, 1, 2, 1],

'priority': ['yes', 'yes', 'no', 'yes', 'no', 'no', 'no', 'yes', 'no', 'no']}

a) Create a DataFrame `df` from this dictionary `data` which has the index `labels`.
b) Display a summary of the basic information about this DataFrame and its data
c) Return the first 3 rows of the DataFrame `df`.
d) Select just the 'animal' and 'age' columns from the DataFrame `df`.
e) Select the data in rows `[3, 4, 8]` *and* in columns `['animal', 'age']`.
f) Select only the rows where the number of visits is greater than 3.
g) Select the rows where the age is missing, i.e. is `NaN`.
h) Select the rows the age is between 2 and 4 (inclusive).
i) Calculate the sum of all visits (the total number of visits).
j) Append a new row 'k' to `df` with your choice of values for each column. Then delete that row to return the original DataFrame.
k) The 'priority' column contains the values 'yes' and 'no'. Replace this column with a column of boolean values: 'yes' should be `True`and 'no' should be `False`.

3. You have a DataFrame df with a column 'A' of integers. For example:

df = pd.DataFrame({'A': [1, 2, 2, 3, 4, 5, 5, 5, 6, 7, 7]})

How do you subtract the row mean from each element in the row?

4.

a) Create a DatetimeIndex that contains each business day of 2015 and use it to index a Series of random numbers. Let's call this Series `s`.
b) Find the sum of the values in `s` for every Wednesday.
c) For each calendar month in `s`, find the mean of values.
d) Create a DateTimeIndex consisting of the third Thursday in each month for the years 2015 and 2016.

5. df = pd.DataFrame({'From_To': ['LoNDon_paris', 'MAdrid_miLAN', 'londON_StockhOlm',
'Budapest_PaRis', 'Brussels_londOn'],
'FlightNumber': [10045, np.nan, 10065, np.nan, 10085],
'RecentDelays': [[23, 47], [], [24, 43, 87], [13], [67, 32]],
'Airline': ['KLM(!)', '<Air France> (12)', '(British Airways. )',
'12. Air France', '"Swiss Air"']})

a) Some values in the Flight Number column are missing. These numbers are meant to increase by 10 with each row so 10055 and 10075 need to be put in place. Fill in these missing numbers and make the column an integer column
b) The From_To column would be better as two separate columns! Split each string on the underscore delimiter _ to give a new temporary Data Frame with the correct values. Assign the correct column names to this temporary Data Frame.
c) n the Airline column, you can see some extra punctuation and symbols have appeared around the airline names. Pull out just the airline name. E.g. `'(British Airways. )'` should become `'British Airways'`.

6.

Given the lists `letters = ['A', 'B', 'C']` and `numbers = list(range(10))`, construct a Multi Index object from the product of the two lists. Use it to index a Series of random numbers. Call this Series `s`.

a) Select the labels `1`, `3` and `6` from the second level of the Multi Indexed Series.
b) Slice the Series `s`; slice up to label 'B' for the first level and from label 5 onwards for the second level.
c) Sum the values in `s` for each label in the first level (you should have Series giving you a total for labels A, B and C).

7. Create a pandas series from: a list, numpy and a dictionary

8. How to combine many series to form a data frame?

9. How to get the items of series A not present in series B?

10. How to get the items not common to both series A and series B?

11. How to get frequency counts of unique items of a series?

12. How to bin a numeric series to 10 groups of equal size?

13. How to stack two series vertically and horizontally ?

14. How to convert the first character of each element in a series to uppercase?

15. How to compute difference of differences between consecutive numbers of a series?

16. Use this dataset

a) Rename the column `Type` as `CarType` in `df` and replace the '.' in column names with '_'.
b) Check if a data frame has any missing values
c) Count the number of missing values in each column
d) Replace missing values of multiple numeric columns with the mean
e) Select a specific column from a data frame as a data frame instead of a series
f) Change the order of columns of a data frame
g) Set the number of rows and columns displayed in the output
h) Format all the values in a data frame as percentages
i) Reverse the rows of a data frame

17. Replace both values in both diagonals of df with 0.

```
df = pd.DataFrame(np.random.randint(1,100, 100).reshape(10, -1))
```

18.Use Iris dataset

a) Set the values of the rows 10 to 29 of the column 'petal_length' to NaN & substitute the NaN values to 1.0
b) delete the column class, Set the first 3 rows as NaN , Delete the rows that have NaN , Reset the index so it begins with 0 again

19. Create the 3 Data Frames based on the following raw data

```
raw_data_1 = {
    'subject_id': ['1', '2', '3', '4', '5'],
    'first_name': ['Alex', 'Amy', 'Allen', 'Alice', 'Ayoung'],
    'last_name': ['Anderson', 'Ackerman', 'Ali', 'Aoni', 'Atiches']}


raw_data_2 = {
    'subject_id': ['4', '5', '6', '7', '8'],
    'first_name': ['Billy', 'Brian', 'Bran', 'Bryce', 'Betty'],
    'last_name': ['Bonder', 'Black', 'Balwner', 'Brice', 'Btisan']}


raw_data_3 = {
    'subject_id': ['1', '2', '3', '4', '5', '7', '8', '9', '10', '11'],
    'test_id': [51, 15, 15, 61, 16, 14, 15, 1, 61, 16]}
```

a) Join the two data frames along rows and assign all_data
b) Join the two data frames along columns and assing to all_data_col
c) Merge all data and data3 along the subject_id value
d) Merge only the data that has the same 'subject_id' on both data1 and data2

e) Merge all values in data1 and data2, with matching records from both sides where available.