Data Science
Lab Exercise (Week 3)

1. Finish Ten Minutes exercise,
   https://pandas.pydata.org/pandas-docs/stable/user_guide/10min.html

2. **Complete the following program**

```
import pandas as pd
data = {'cities' : ['lahore','karachi',], 'provinces' : ['punjab','sindh']}

# store data as DataFrame object. Assign object name as frame1
frame1 = _____

# print frame

_____

data2 = {"cities": ["islamabad","karachi","peshawar","quetta"],
  "provinces": ["capital","sindh", "KPK","Balochistan"]}

# store data as DataFrame object. Assign object name as frame2

_____

# combine both objects frame1 and frame2; without any duplicate rows and re-arrange all
indexes

frame3 = ………………………………………    # combine frame1 and frame2
frame3 = …………………………………………   # remove duplicates rows
frame3 = …………………………………………….# sort based on provinces
frame3 = …………………………………………….# re-arrange all indexes
………………………………………………………… # print frame3
```
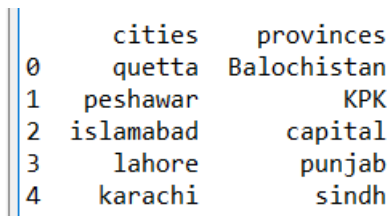
```
         cities    provinces
0        quetta   Balochistan
1       peshawar          KPK
2      islamabad      capital
3         lahore       punjab
4        karachi        sindh
```
Figure1: Screen shot of Final Output for Q2.

3. Consider the following table

Name	Field	Age	Marks
	C		-90
Ali	E		60
Ahmed	E		-10
Nida	C		70
	C		75

Perform following data cleansing operation on the given data.

      i.   Drop column **Age** as it does not contain any value

      ii.   All empty strings in the **Name** column should be replaced by "---"

      iii.   In the **Field** column replace "C" with 0 and "E" with 1. The column must contain only numeric values after this operation

      iv.   Negative values are not permitted in **Marks** column. The invalid value in **Marks** column should be replaced with the average of all valid values in the same column

4. Finish Exploratory Data Analysis with Pandas from the following link
https://www.kaggle.com/kashnitsky/topic-1-exploratory-data-analysis-with-pandas
https://www.kaggle.com/ekami66/detailed-exploratory-data-analysis-with-python