CS125 Introduction to Data Analysis for Social Sciences
Lab Guide 09

Objectives: Pandas – data manipulation (loc, iloc, square braces, boolean expressions)

1. Download the file `clinic_data.xlsx`, and create a Python script that does the following:
   a. Import the data in the file into a DataFrame.
   b. Display the names of all doctors.
   c. From the 1st and 4th columns, display the 3rd to the 5th rows (inclusive)
   d. For all records, display the year, months and number of patients.
   e. Display the records for visits in November.
   f. Display the record with the maximum patients.
      (hint: to find the max, `df.col_index.max()`)
   g. Display records from September of February.
   h. Display records after 2014 with fewer than 300 patients.
   i. Display records from September or February with over 250 patients.
   j. Set the first column in the DataFrame as the index.
      (Hint: `df.set_index(['column_label'], inplace = True)`) By setting the `inplace` attribute to `True`, it will update the original DataFrame.
   k. Use loc[] to display the year and patients for doctors Benitz and Carli.

Sample Run:

```
a) Data:
        doctor   year      month  patients
0       Benitz   2012   November       172
1        Modin   2013     August       272
2       Leabow   2014   November       203
3   Vandergard   2014  September       259
4       Largay   2014     August       293
5        Jonas   2015   February       233
6        Carli   2017   February       301
7         Palm   2019      April       396
8      Dukette   2018    October       407
9        Modin   2010   November       390

b) Doctors Names:
0         Benitz
1          Modin
2         Leabow
3      Vandergard
4         Largay
5          Jonas
6          Carli
7           Palm
8        Dukette
9          Modin
Name: doctor, dtype: object
```

c) From the 1st and 4th columns the 3rd to the 5th rows (inclusive):

```
      doctor  patients
2     Leabow       203
3  Vandergard       259
4     Largay       293
```

d) Year/Month/Patients:

```
   year        month  patients
0  2012   November       172
1  2013     August       272
2  2014   November       203
3  2014  September       259
4  2014     August       293
5  2015   February       233
6  2017   February       301
7  2019      April       396
8  2018    October       407
9  2010   November       390
```

e) Records from November:

```
   doctor  year       month  patients
0  Benitz  2012  November       172
2  Leabow  2014  November       203
9   Modin  2010  November       390
```

f) Record with max patients:

```
    doctor  year     month  patients
8  Dukette  2018  October       407
```

g) Records from September or February:

```
       doctor  year       month  patients
3  Vandergard  2014  September       259
5       Jonas  2015   February       233
6       Carli  2017   February       301
```

h) Records after 2014 with fewer than 300 patients:

```
  doctor  year      month  patients
5  Jonas  2015  February       233
```

i) Records from September or February with over 250 patients:

```
       doctor  year       month  patients
3  Vandergard  2014  September       259
6       Carli  2017   February       301
```

k) Year and Patients for Benitz and Carli:

```
        year  patients
doctor
Benitz  2012       172
Carli   2017       301
```