KOÇ ÜNİVERSİTESİ

INTRODUCTION TO COMPUTATIONAL SCIENCE
CMSE 501

Pandas Dataframe Exercise

You are given a csv file in the following format including data on life expectancy and GDP per capita for each country during the years 1952-2007.

	country	year	pop	continent	lifeExp	gdpPercap
0	Afghanistan	1952	8425333.0	Asia	28.801	779.445314
1	Afghanistan	1957	9240934.0	Asia	30.332	820.853030
2	Afghanistan	1962	10267083.0	Asia	31.997	853.100710
3	Afghanistan	1967	11537966.0	Asia	34.020	836.197138
4	Afghanistan	1972	13079460.0	Asia	36.088	739.981106

…

You are asked to prepare the following requests using pandas dataframe:

1. Prepare a subset of the data frame based on values of year column; for year 2002
   a) Filter rows for year 2002 using a boolean variable
   b) Filter rows for year 2002 using a boolean expression
2. Select rows whose column value is in list, e.g. values for year [1952, 2007]
3. Select rows whose column value is not among the list of values provided, e.g. rows with a value for continent not among the values provided by the list ['Asia','Africa', 'Americas', 'Europe']
4. Select rows by combining the above two conditions (2) and (3) to get Oceania data from years 1952 and 2002

Data file : LifeExp.csv (Located in Blackboard Exercises folder)