

Class 1 Homework Exercises

1. Explain in your own words what this piece of code is doing:

```
df = read_csv("test_csv.csv")
df2 = df.loc['row1':'row6', :]
df2.head(5)
```

The code reads in a csv file called test_csv as a pandas dataframe and assigns it to the variable df. The it takes a subset of that dataframe between row1 and row6, preserving all of the columns, and assigns that subset as a new dataframe to the variable df2. Then it displays the first 5 rows of df2.

2. What is the index column of the above df? What parameter would we use to specify the column of index 5 as the index column for the dataframe?

The dataframe df will have a numerical index created by pandas, starting at 0. To assign column 5 as the index column, we load the dataframe with this command:

```
df = read_csv("test_csv.csv", index_col=5)
```

3. Write the code the get the minimum value of a row “year” in a dataframe df_years

```
df_years.loc["year"].min()
```

4. If you are about to write some code using the pandas library, what is the first line of code you have to enter in your program?

```
import pandas as pd
```

5. Below is a table containing the population of major cities in millions by year. Write the code to print which cities in the following table had a population greater than 20 million in the year 2010.

Note: I have made up these numbers except for 2020!

	London	Paris	New York	Tokyo
2020	9.0	10.9	18.9	37.5
2010	8.9	10.4	19.2	37.4
2000	8.9	10.5	19.1	37.1
1990	8.8	10.4	19.0	36.5

```
countries = df.columns.values
```

```
for x in countries:
```

```
    min = df[x].min()
```