1. **Apply**

Special thanks to: https://github.com/justmarkham for sharing the dataset and materials
United States - Crime Rates - 1960 - 2014

Step 1. Import the necessary libraries

Step 2. Import the dataset from this https://raw.githubusercontent.com/guipsamora/pandas_exercises/master/04_Apply/US_Crime_Rates/US_Crime_Rates_1960_2014.csv

Step 3. Assign it to a variable called crime.

Step 4. What is the data type of the columns?

Have you noticed that the type of Year is int64. But pandas has a different type to work with Time Series. Let's see it now.

Step 5. Convert the type of the column Year to datetime64

Step 6. Set the Year column as the index of the dataframe

Step 7. Delete the Total column

Step 8. Group the year by decades and sum the values

Pay attention to the Population column number, summing this column is a mistake.

Step 9. What is the most dangerous decade to live in the US?

2. **Stats**

We are going to use a subset of [US Baby Names](https://www.kaggle.com/kaggle/us-baby-names) from Kaggle
In the file it will be names from 2004 until 2014

Step 1. Import the necessary libraries

Step 2. Download and extract the data from this link https://drive.google.com/open?id=1At_YCsguCIEqe3l-gSPPquaOORbGLKmj

Step 3. Assign it to a variable called baby_names.

Step 4. See the first 10 entries

0' and 'Id'

Step 6. Is there more male or female names in the dataset?

Step 7. Group the dataset by name and assign to names

Step 8. How many different names exist in the dataset?

Step 9. What is the name with most occurrences?

Step 10. How many different names have the least occurrences?

Step 11. What is the median name occurrence?

Step 12. What is the standard deviation of names?

Step 13. Get a summary with the mean, min, max, std and quartiles.

3. **Visualization**

Step 1. Import the necessary libraries

Step 2. Import the dataset from this https://raw.githubusercontent.com/justmarkham/DAT8/master/data/chipotle.tsv

Step 3. Assign it to a variable called chipo.

Step 4. See the first 10 entries

Step 5. Create a histogram of the top 5 items bought

Step 6. Create a scatterplot with the number of items ordered per order price,

Price should be in the X-axis and Items ordered in the Y-axis

Create a question and a graph to answer your own question.

4. **Creating Series and DataFrames**

This time you will create the data

Step 1. Import the necessary libraries

Step 2. Create a data dictionary that looks like the DataFrame below

	Evolution	HP	Name	Pokedex	Type
0	Ivysaur	45	Bulbasaur	yes	grass
1	Charmeleon	39	Charmander	no	fire
2	Wartortle	44	Squirtle	yes	water
3	Metapod	45	Caterpie	no	bug

Step 3. Assign it to a variable called

Step 4. Oops...it seems the DataFrame columns are in alphabetical order. Place the order of the columns as name, type, hp, evolution, pokedex

Step 5. Add another column called place, and insert what you have in mind.

Step 6. Present the data type of each column

Create your own question and answer it.

5 **Time_Series**

Step 1. Import the necessary libraries

Step 2. Import the dataset from this

https://raw.githubusercontent.com/datasets/investor-flow-of-funds-us/master/data/weekly.csv

Step 3. Assign it to a variable called invest

Step 4. What is the frequency of the dataset? (i.e., what's the time frequency?)

Step 5. Set the column Date as the index.

Step 6. What is the data type of the index?

Step 7. Set the index to a DatetimeIndex type

Step 8. Change the frequency to monthly, sum the values and assign it to monthly.

Step 9. You will notice that it filled the dataFrame with months that don't have any data with NaN. Let's drop these rows.

Step 10. Good, now we have the monthly data. Now change the frequency to year.

Create your own question and answer it.