**Overview**

Imagine you are working for a company that is developing a night lamp for hotels that is connected to the hotel’s smart network. Guests at the hotel can control the night lamp in their room using their smart TV remote control. Your company is considering expanding into the European market but is not sure which national market to launch in first. You are tasked with producing a data set that can be used to make an informed decision about which national market is the best fit.

### **Task**

Your task consists of two parts, the first part is to combine two public data sets into a single CSV file that contains 3 columns: “Country Code”, “Percentage of individuals” and “Number of Bed-places”. The “Country Code” column should be distinct and should not contain missing data.

The second part of your task is to somehow visualize the resulting data set, you may visualize the data in whatever way you see fit, both plots and tables are fine. Please produce at minimum 1 and at maximum 3 visualizations in total.

### **Datasets**

The raw data comes from two separate datasets in .tsv.gz (tab-separated, gzip-compressed) file format:

**TOUR\_CAP\_NAT**

Description: Number of establishments, bedrooms and bed-places

URL:<https://data.europa.eu/data/datasets/ndozfy78qqwgcenbwpjxvg?locale=en>

Vocabulary:

| accomunit | Value type (e.g. bedrooms, bed-places) |
| --- | --- |
| unit | Value unit (e.g. absolute, percentage) |
| nace\_r2 | Classification of economic activity |
| geo | Country Code |
| TIME\_PERIOD | Year for which the data was collected |
| OBS\_VALUE | Number of bedrooms, bedplaces or establishments |

Columns to use: “accomunit,unit,nace\_r2,geo” and “2016 “

Rows to use: You should only use rows where:

· “accomunit” is “BEDPL”

· “unit” is “NR”

· “nace\_r2” is “I551”

· “TIME\_PERDIO” is “2016”

**isoc\_ci\_dev\_i**

Description: devices used to access the internet

URL:<https://data.europa.eu/data/datasets/davj8w8vlmqipfyvgobniw?locale=en>

Vocabulary:

| ind\_type | Population segments, e.g. students, males, females etc. |
| --- | --- |
| indic\_is | Device type identifier, eg   * I\_IUG\_DKPC – desktop PC * I\_IUG\_LPC – laptop PC * I\_IUG\_MD – mobile device * I\_IUG\_TV – TV |
| unit | The data unit, percentage of individuals. |
| geo | Country Code |
| TIME\_PERIOD | Year for which the data was collected |
| OBS\_VALUE | Percentage of individuals in a country |
| OBS\_FLAG | Quality of the data in OBS\_VALUE |

Rows to use:

· “ind\_type” is “IND\_TOTAL”

· “indic\_is” is “I\_IUG\_TV”

· “unit” is “PC\_IND”

· “TIME\_PERIOD” is “2016”

### **Special values**

● The value “: ” signifies missing data.

● Rows where OBS\_FLAG contains “u” or “bu” are unreliable and should be considered as missing data.

● Country codes “EA”, "EU27\_2007", "EU27\_2020", "EU28" should be treated the same as missing data and should be ignored.[2]

### **Submission**

Your solution must be written in Python and should include

· The merged CSV file, with columns “Country Code”, “Percentage of individuals” and “Number of Bed-places”.

· Plots and/or tables for visualizing the data as separate files.

· The code you used to produce the merged CSV and plots/tables.

You can either email your solution as a zip file directly to [ailysrecruit@ailys.ai](mailto:ailysrecruit@ailys.ai) or upload your solution to github, gitlab or bitbucket and email the link to your repository to ailysrecruit@ailys.ai

### **Notes**

· If the instructions are unclear do not hesitate to ask us for clarification!

· There is no need to make the solution advanced. Doing basic things should be enough.