1711-SMU-DSiP_KRK

Advanced Python

Data Analysis Mini-project using Pandas & Matplotlib

Deadline : Sunday, January 31st, 2021
Submission : a concise report of the project and the python sources. Email them to the address **hayfa.nakouri@esen.tn**

The dataset "amazon.csv" is about the forest fires in Brazil, which houses the largest rainforest on Earth – Amazon.

The dataset has 5 columns:
year: is the year when the forest fire happened;
state: is the Brazilian state;
month: is the month when the forest fire happened;
number: is the number of forest fires reported;
date: is the date when the forest fire was reported

1. Read the data and view it.
Important : add the parameter "**encoding='latin1'**" to **pd.read_cvs()** because the data is in Portuguese. ('á' and 'ç' will generate an error).

2. The months in the column "month" are in Portuguese. Change them to English.

3. Missing values in the state column:
a- Randomly delete 200 values from the state column;
b- Check for the missing values in this column;
c- Fill each missing value with that of the previous (or next) row.

4. Data reduction:
We would like to visualise the number of forest fires over the years, over the months. To this, we need to be able to identify smaller pieces of the bigger picture.

Let us drop rows from the dataset that are not contributing to the number of forest fires. Therefore, any row with number column value as 0, must be dropped. We first convert the 0s to NaN and then drop rows with NaN in the specific column number.

5. We want to group data per month and show the total number of fires for each month. The output should be in a Pandas series.

You will notice that the output orders the months in alphabetical order. Transform it to a DataFrame and get it back to the monthly order. The final DataFrame should look like the right table :

	Month	Number
0	January	52587.0
1	February	30952.0
2	March	35118.0
3	April	28364.0
4	May	46083.0
5	June	111405.0
6	July	217620.0
7	August	740841.0
8	September	1015925.0
9	October	629665.0
10	November	312326.0
11	December	152596.0

6. Finally, bar graph the DataFrame found in Question five.
The bar graph should illustrate the number of forest fires per month.