DA25062021

June 25, 2021

1 Data Analytics June 25 2021

1.0.1 Surname

Write your Surname here

1.1 Name

Write your Name here

1.1.1 Matricola ID

Write your Matricola here

[2]:
```
import pandas as pd
```

The file `Views.csv` contains users views of 4 different channels of a streaming service. The channels are `['Sport', 'Movies', 'News', 'Fiction']`. In each row are reported: the date of the view (`Date`), the channel (`Channel`), the viewing time in minutes (`Duration`) and the user id (`User`, `integer`).

The file `Calendar.csv` contains all days of year 2020 along with the name of each day.

1. Read the file `Views.csv` into a convenient dataframe.
2. If the dataframe contains `NaN` values in the `Duration` field, replace it with 0, if it contains `NaN` in other fields drop the corresponding row.
3. Extract the total number of users.
4.
    1. Create dataframes containing:
        1. The average and total viewing time for each channel in year 2020.
        2. The average and total viewing time for each day of the week.
        3. [**Bonus**] Sort the dataframes of point b. by weekdays, i.e., Mon, Tue, Wed, Thu, Fri, Sat, Sun.
    2. Graph the distribution of the viewing times for each channel.
5. Create a multi-indexed dataframe containing the viewing time for each day of the week (first level index) and for each channel (second level index). What is the most watched channel over the weekend?
6. **Bonus question**

Assuming viewing times have the following costs per channel:

- `'Sport'`: 0.35 Euro/min for duration greater than 10 minutes, 0 otherwise.
- `'Movies'`: 0.20 Euro/min for duration greater than 15 minutes, 0 otherwise.
- `'News'`: 0.10 Euro/min for duration greater than 5 minutes, 0 otherwise.
- `'Fiction'`: 0.25 Euro/min for duration greater than 10 minutes, 0 otherwise.

Calculate the bill for each user.

[ ]: