

Python in engineering calculations

Project's report

NOTE: The script was written and tested in Windows. Some functions might not work properly on other systems such as Linux.

1. Used packages

All necessary packages are listed in *env.yml* file that is needed for creating new environment where the script can be executed. Command: `conda env create -f env.yml`

2. Directory structure

In order to execute the script correctly there must be some directory structure:

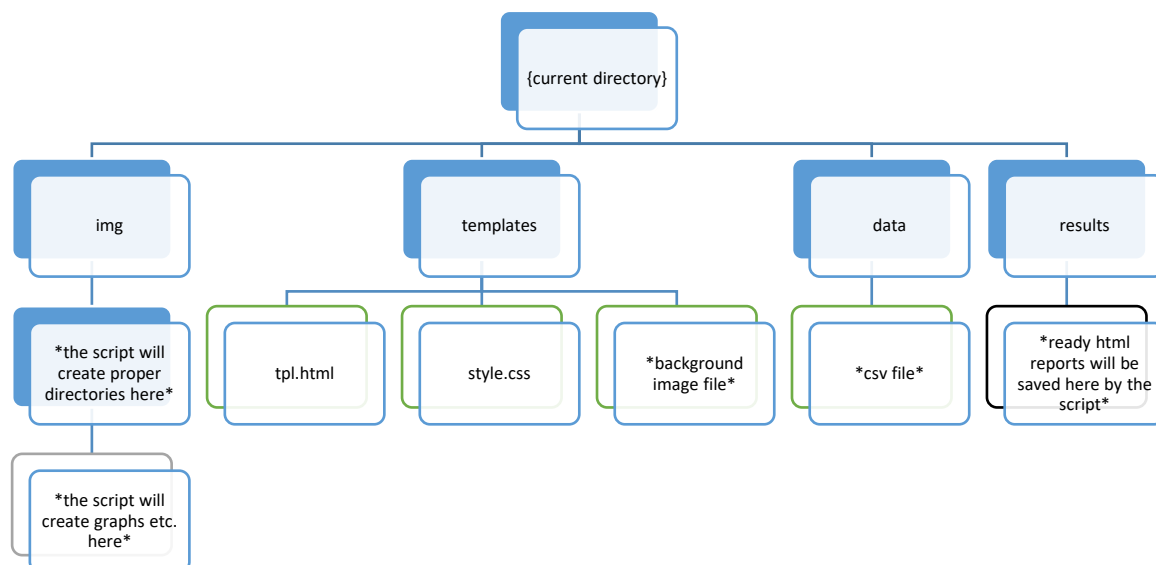


Figure 1. Required directory structure.

3. Preparation of downloaded file

In order to perform data analysis, original file (downloaded here: <https://www.kaggle.com/yamaerenay/spotify-dataset-19212020-160k-tracks>) was slightly modified – for example: there were made changes to durations of the songs from milliseconds to seconds or dropping (irrelevant) column with IDs from Spotify. Those attempts are shown below.

```
In [8]: dlugosc = []
        for i in range(len(df.duration_ms)):
            cosik = df.iloc[i]['duration_ms']
            czas = datetime.timedelta(milliseconds=cosik.item())
            czas = czas - datetime.timedelta(microseconds=czas.microseconds)
            dlugosc.append(sum(x * int(t) for x, t in zip([3600, 60, 1], (str(czas).split(":")))))

        dlugosc
        df['length'] = dlugosc
        df = df.drop(columns=['duration_ms', 'id'])
        df
```

Figure 2. Fragment of the code (from *jupyter notebook* file) in which additional columns were dropped.

Such prepared data has been saved to another file which will be later used in the script.

4. Script description

The script has two modes – the first that allows user to compare two decades and the second that allows to analyse whole time period (1921 – 2020). Both modes include insight into full range of factors that are described further in report page. In addition, the second mode shows table of correlation coefficients between all categories in order to examine whether some variables might be connected with each other.

Basically the script requires two arguments that will determine which mode should be executed. The third argument is optional and it contains the name of the output report. Below there are listed examples of commands:

- The first mode – 2 decades comparison :
python make_report.py 70s 10s
- The second mode – years from 1921 to 2020:
python make_report.py all decades
- Changing name of the report (2 ways):
python make_report.py 70s 10s --name report
or shorter:
python make_report.py 70s 10s -n report

The ready report will include any necessary images and information.