Finnish Transport Agency Analysis

Overview

Digitraffic is a service of Traffic Management Finland, through which up-to-date traffic information is available on Finnish road network, rail traffic and maritime traffic. Using their open traffic data through open interface we perform following tasks.

1) Collect and normalize the July 2020 month data for train number 4 and store the file in csv format.
2) Calculate the average actual arrival time at the final destination of the train number 4 during the month of July 2020.

Task1 steps:

1) Create the list to store each days of July 2020 month API for the train number 4.
2) Fetch each day's data using API from the Finnish Transport Agency.
3) Fetched data is in json format, to further analysis we have flatten the data using json_normlize from pandas' package.
4) Store the flatten data for each day in dataframe format into list.
5) Combine each dataframe from the list into one final dataframe.
6) Store the final dataframe on local disk as csv format.

Task2 steps:

1) Each day train number 4 reaches at the final destination station with station UIC code 1 and station short code HKI.
2) We use final dataframe and filter the dataframe with station UIC code 1 and type ARRIVAL and get the actual time column which has actual arrival date and time of the train at the final destination.
3) We extract the time from actual arrival time column.
4) Convert each time into seconds and take average for whole month into seconds
5) Convert those seconds back to actual time.
6) Display average actual arrival time of train number 4 for whole month of July-2020

Research:

During the task, we have found that below days data is missing from Finnish Transport Agency. we completed the task using rest available data for the July 2020 month.

05-07-2020
12-07-2020
19-07-2020
26-07-2020

Improvement:

Further, we can add more columns such as categoryCode, categoryCodeId, detailedCategoryCode, detailedCategoryCodeId etc. store into causes node of json data. We could add those data by flattening the json data with metadata as causes which will fetch single row of each json data and store it into another dataframe. finally, merge this dataframe with final dataframe.

At this moment, those columns do not require to complete the task, so we skip it. But for feature analysis we can use it.

Conclusion:

Using the Finnish Transport Agency open interface, we have fetched the json data of train number 4 and store it in local disk drive with require format. Also, we have calculated the Actual arrival time of train number 4 at destination station for the month of July 2020.