Data Engineer Excercise Assignment

Most of our personnel in Finland work in Tampere (TPE) and Helsinki (HKI) and have to travel this distance by train in the name of customer experience. In addition to high-quality customer service, our culture includes caring about employees spare time activities.

Jaana's son has a basketball game in the center of Tampere at 4:00 pm, where he has promised to come and cheer on him. To get to the basketball game from Helsinki, Jaana has booked a ticket to the pendolino nr. S45.

Predict if Jaana can arrive next weeks Wednesday exactly at 4 pm to the event in Tampere with the Pendolino S45 (transfer from the station to the event takes 2 minutes) or should he leave the work on an earlier train?

Take advantage of historical data for the prediction from the actual arrival times of the trains. Please retrieve data from the Open Data Portal of the Finnish Transport Agency.

1. The Open Data Portal of the Finnish Traffic Agency for railway traffic: https://rata.digitraffic.fi

a. Railway traffic history: https://rata.digitraffic.fi/#junien-tiedot-trains

Translation: https://translate.google.com/translate?hl=fi&sl=fi&tl=en&u=https%3A%2F%2Frata.digitraffic.fi%2F%23junien-tiedot-trains&sandbox=1

b. Example API url for one day's data: https://rata.digitraffic.fi/api/v1/trains/2017-01-01/45

c. Attached is an obsolete CSV file that you can take advantage of if the API interface causes too much grey hair

(Note, all times from the API is UTC time)

2. After you retrieved the data you need, load it to a database
3. Make a forecast of the arrival time of the train to Tampere Station
4. Provide the justification for the forecast you have provided in an easily understandable format, for example through visualizations.
5. Create an application that provides an interface that tells what time the S45 will be next Wednesday at Tampere
6. If there is time, write down in a few sentences how otherwise the open data provided by the Finnish Transport Agency could be used to promote to bring business value: what are the possibilities to use the data, what is good about it and what could be the problems?

Take advantage of the tools that you are most effective when doing the job. If you have extra time left, document your work and tell how your solution could have been further improved.