Dear BMW applicant,

Thanks again for your application. To test your ability to work with multiple data sources we would like you to answer the tasks below. We expect to receive your commented code as well as a short text description on your observations about the problems, the results and why you did the implementation the way you did it.

Please solve the problems in Python if possible.

If you do not have time to solve all four tasks, please focus on the first and the second task.

Good luck!

Your BMW Research Team

Task 1:

This task is aiming to provide a binary classification of the column “Type”. The dataset is provided into two different tables with unique identifier of column “ID”. Tip: This column (ID) can be used to match the two tables.

Task 2:

The file “Parts.csv” contains descriptions of some fictitious parts. Your goal is to find 5 alternative parts to each provided fictitious part in the dataset based on their similarity. First provide descriptive analysis of the data and highlight 2-3 findings and difficulties of the data that we provided and describe how you would handle this. Continue to implement a model that is finding the similar fictitious parts based on the column “DESCRIPTION”. Please give details of your model and why you choose this. Once you finished your implementation of your chosen model, please think about how you would change or addition your code, when you have more entries and would consider the additional attributes.

Task 3:

Write a function that takes as input two timestamps of the form 2017/05/13 12:00 and calculates their differences in hours. Please only return the full hour difference and round the results. E.g. 2022/02/15 00:05 and 2022/02/15 01:00 would return 1 hour.

Task 4:

Expand the above function to only count the time difference between 09:00 – 17:00 and only on weekdays.