**File Descriptions:**

1. Company\_all\_tables(unstructured): This is an excel file with multiple sheets, each sheet contains a particular table in it with specific table\_id , These are the abby screen reader converted output from pdf into an excel format, these contains different tables with different structure and time series(basically years and its values) but in an unstructured manner with special characters and all.
2. Company\_quant\_file(structured): This is also an excel file but in structured format, you can see different column names present, these contains the similar table\_id , names and values from the “company\_all\_table(unstructured)” but in a structured manner, with differentiation of dim id’s(dim\_4\_\_name, dim\_3\_\_name, dim\_2\_\_id and dim\_1\_\_name and metric\_name) and then the same value under the same time series/years(2022/2021/2020).

**Code already working on :**

1. My code is already cleaning the special characters from the “Company\_all\_tables(unstructured)”.
2. It is comparing simple table\_id with just metric\_id and values of “Company\_all\_tables(unstructured)” with “Company\_quant\_file(structured)”

**Required Criteria From Freelancer:**

1. It should filter the table\_id of **“**Company\_quant\_file(structured)” on the basis of **“**dim\_ name” and extract different table with different criteria. Suppose a table\_id only has metric\_ name present so consider it one table type criteria, now if a table\_id contains “metric\_ name” and only “dim\_1\_ name” consider it as another table type, and if a table\_id contains “metric\_id” ,”dim\_1\_ name” and “dim\_2\_ name” consider it as another and so on.
2. Compare the different filtered table\_id of “Company\_quant\_file(structured)” with the same table\_id of “Company\_all\_tables(unstructured)”, match the metric\_name and dim\_name of the “Company\_quant\_file(structured)” with the “Company\_all\_tables(unstructured)” and the value in time should be exact same in front of metric name for both the file
3. Compare all the tables and at last give an output if there is value error or the exact metric in both the file does not have same value in same time series infront of similar metric name.