

## Spark SQL hiring case study

The files needed for this hiring case study are:

- SE\_Sample\_data.csv
- R3F015.csv
- DealerCurrency.csv
- FORECAST\_TOOL (Test Case).ipynb

Objective:

We have 3 sources of information, SE\_Sample\_data.csv, R3F015.csv, DealerCurrency.csv. The business needs to combine those files and to change the format of the final table before storing it in a SQL database, where the DIM analysts will use for further analysis and creating visualizations.

The tasks required in the test case are the following:

1. Change the code provided as needed to run it in your environment. The code has pre-written directory address, you need to change the code for your own storage space.
2. Find and fix bugs in the FORECAST\_TOOL (Test Case).ipynb file provided. The code has some bugs that will show up when you try to run it. We recommend making a note of the bugs found and how did you enhance the code.
3. Once the first part of the code is working correctly. The next task is to put the information in tidy format. i.e. there is a need to unpivot the following columns:
  - SalesPMs as (SalesMonthU, SalesUnits),
  - SalesPMs \* Price \* CurrencyRate as SalesSEK,
  - ForecastPMs AS (ForecastMonthU, ForecastUnits),
  - ForecastPMs \* Price \* CurrencyRate as ForecastSEK,
  - ForecastMethod as (ForecastMethodMonthU, ForecastMethod),
  - NoOfPicksPerMs as (NoOfPicksMonth, NoOfPicks)

For example: Originally, the ForecastMethod is stored in the dataset, on the following columns:

```
tv_R3F002.ForecastMethodM00,  
tv_R3F002.ForecastMethodM01,  
tv_R3F002.ForecastMethodM02,  
tv_R3F002.ForecastMethodM03,  
tv_R3F002.ForecastMethodM04,  
tv_R3F002.ForecastMethodM05,  
tv_R3F002.ForecastMethodM06,  
tv_R3F002.ForecastMethodM07,  
tv_R3F002.ForecastMethodM08,
```

tv\_R3F002.ForecastMethodM09,  
tv\_R3F002.ForecastMethodM10,  
tv\_R3F002.ForecastMethodM11,  
tv\_R3F002.ForecastMethodM12

The idea is to replace those 13 columns with 2 columns, one column for the Forecast method (the value in column tv\_R3F002.ForecastMethodMX), and another column for the Forecast method month (for example 12 for tv\_R3F002.ForecastMethodM12).

Similar for the other columns (ForecastPMs, NoOfPicksPerMs , ...).

4. Additional task and optional. Add comments to the code and enhance it for better performance.

We recommend that you use Microsoft Azure Databricks (optional). You can get a free account here:

[https://databricks.com/try-databricks?utm\\_medium=paid+search&utm\\_source=google&utm\\_campaign=14925725269&utm\\_adgroup=133613194972&utm\\_content=trial&utm\\_offer=try-databricks&utm\\_ad=563736643330&utm\\_term=databricks%20inc&gclid=EAlaIQobChMI29XQ26mT9wIVGkiRBR1D8QfCEAAAYASAAEgI9RvD\\_BwE](https://databricks.com/try-databricks?utm_medium=paid+search&utm_source=google&utm_campaign=14925725269&utm_adgroup=133613194972&utm_content=trial&utm_offer=try-databricks&utm_ad=563736643330&utm_term=databricks%20inc&gclid=EAlaIQobChMI29XQ26mT9wIVGkiRBR1D8QfCEAAAYASAAEgI9RvD_BwE)

During the meeting you will be asked to go through your code and explain the changes that you have done, and the new code written. Also, to summarize what the code is doing.