Dear candidate,

Thank you for your interest in joining Conrad. We appreciate your interest in our company and your participation in our recruitment process. Following is the case study for the Data Science position.

Imagine that the purchasing department wants to anticipate future products sales at Conrad. This would help optimize the stock and save costs while guaranteeing the product's availability for our customers. For that, we would like to implement a Machine Learning model predicting future products sales.

The available data would look as follows:

	productsGroup_key	date_key	quantity Sales
0	1	20190902	26784.0
1	1	20190903	7432.0
2	1	20190904	1424.0
3	1	20190905	608.0
4	1	20190906	776.0
5	1	20190907	808.0
6	1	20190908	104.0
7	1	20190909	632.0
8	1	20190910	7256.0
9	1	20190911	1176.0

Tasks:
1. In that case, what would be the business KPIs for evaluating the impact of the new ML model?
2. Which loss functions could be used for the model evaluation? Mention the reason for that?
3. What benefit would you see when using a probabilistic approach vs a single point forecasting approach for tackling this use case?
4. Imagine that the sales results are influenced by the Covid-19 spread as well as the inflation situation. How would you deal with this impact on the data?
5. Implement a Machine Learning model of your choice to estimate the uncertainty distribution and the accuracy of Conrad sales.

6. Provide the submission.csv file after filling out the fields containing zero values with the proper model output

	productsGroup_key	date_key	quantitySales	q_0850	q_0900	q_0920	q_0950	q_0990
818	1	20211128	0	0	0	0	0	0
819	1	20211129	0	0	0	0	0	0
820	1	20211130	0	0	0	0	0	0
821	1	20211201	0	0	0	0	0	0
822	1	20211202	0	0	0	0	0	0
823	1	20211203	0	0	0	0	0	0
824	1	20211204	0	0	0	0	0	0
825	1	20211205	0	0	0	0	0	0
826	1	20211206	0	0	0	0	0	0
827	1	20211207	0	0	0	0	0	0

Provided files:
- train.csv: training data
- submission.csv: results submission file

Description of the provided data fields

- productsGroup_key - products group which consists of different articles
- date_key - date in format yyyymmdd
- quantitySales - sum sales of all articles in a products group
- q_* - probability target variables (in the submission.csv file)

Note: the provided data should not be made publicly available

We are looking forward to getting your solution to the task in the form of your choice (Github repository / Jupyter notebook(s), …) plus the submission.csv file, and discussing it.
Good luck.

Best regards,
Conrad Data Science Team