



## Advanced Data Analysis Project

You have been tasked to perform an analysis by your Data Science Manager, Jack Analytisson. Jack is interested in identifying variables for his ML model. The machine learning model he wants to build is a demand forecasting model. The model is going to be forecasting the sales over the next period. He has provided you the database *WorldWideImporters* of the company with all relevant data in the organization. Unfortunately, he does not know what data might be useful or interesting. It is your task to explore the database to:

- Identify potential useful tables/data for his model
- Do statistical analysis on the selected data above
- Visualize your findings

Jack has advised you, that a lot of effort will be required to carry out the analysis therefore has given you plenty of time to prepare. Jack would be delighted if you could build an initial regression model (in Python or Excel). Think outside the box and provide your ideas of insights as well on top of what Jack has asked you. Base your arguments on the data. Provide a short explanation of your thinking why your suggestions can bring business results. Add comments in your code explaining what you are doing.

You are expected to use the data in the *WorldWideImporters* database we have provided you, using the credentials assigned to you. You can use SQL to extract the data. Excel to carry out your initial analysis. However, Jack requested that visualization and further analysis should be done in python (ideally on Jupyter notebook, but is ok with .py scripts too).

You can work in groups if you want to, but you are expected to submit a .zip file **individually** that contains ALL of your own work:

- SQL scripts

- Excel drafts / experiments
- Python code
- Presentation of your findings and recommendations

Submit your work findings via e-mail to [REDACTED] no later than **Monday 13<sup>th</sup> June 23:59**.