Argus Media: Modelling Exercise using GAMLSS

At Argus we are interested in the way you approach a data science problem and how you present and/or support your reasons with data; the goal of this test is to this. Using the oil data from the GAMLSS package carry out the following:

1. Conduct a brief preliminary exploratory data analysis (EDA) of the dataset to discover any patterns, spot potential anomalies, test initial hypotheses and check assumptions with the help of summary statistics and graphical representations. Briefly discuss your findings. Please note that the variable "OILPRICE" is the response variable while all the remaining variables can be used as your covariates.
2. Given your EDA, do you need to need to consider transformations of your variables? Do you need to create additional features which could lead to more predictive power. Briefly discuss your findings.
3. Develop a simple initial model for predicting the one day ahead probability distribution of the oil price using the GAMLSS framework. Consider splitting your data into training and testing to evaluate the accuracy of your model and briefly discuss whether the model is accurately capturing the one day ahead distribution or if further considerations are needed (note you need to evaluate a distribution forecast and not a point forecast). Please note that at this stage there is no need to have an accurate model; we want you to determine a baseline.
4. Try now to perform feature selection to find covariates which improves your initial model. What metric will you use to determine which features are better and thus be included in the model? Evaluate the accuracy of your new model using visuals, statistical tests etc.
5. Given your new model, are there any further considerations that could improve the model. Could you tune or change parameters in the model to improve the predictive performance. Briefly discuss your findings.

The preferred way to report your findings is a pdf/HTML file generated using R markdown.

You can use the code below to install the **gamlss** package in R and extract the oil data.

```
 # install the gamlss package
install.packages(c("gamlss","gamlss.add","gamlss.dist"))
library(gamlss)
library(gamlss.add)
library(gamlss.dist)
# extract the oil dataset
data(oil)
```

Deliverable: R Markdown file **AND** the corresponding HTML/PDF report
Notes: We are mostly interested in your approach to model development. You need to demonstrate the steps you follow when developing a predictive model. For example, data exploration, feature engineering, feature selection, model testing/validation, etc. Focus on demonstrating the modelling steps you would follow.