Technical Assessment for Data Analyst

This technical assessment consists of

- Dataset: dataset.csv
- Assignment(s): listed in this document

Please submit your solution(s) for the following assignment(s):

1. Data Cleansing & Processing

Build a solution to automate below processes for the dataset

- Transform numeric column type into text
- Remove duplicated row(s) and keep only row with latest [YearBuilt]
- Automatically detect and remove outliers for column [SalePrice]
- Replace all NA with blank and handle missing values (blank): for numeric column, fill up with average value; for categorical column, fill up with most-frequent value
- Check if one [LotArea] is having multiple [LotFrontage]: if so, keep the row with largest [LotFrontage] for that [LotArea] only
- Check if [SalePrice] is trending upwards over the years [YrSold] for the same [MSSubClass]: for those not, display the [MSSubClass] value and indicate the year [YrSold] when [SalePrice] trend is reversed

Requirement

- Use Python to develop the solution and pack it up to an importable format which is allowable to be utilized as a library
- The solution should cater for exceptions and be informative on such as error message display, progress-bar display and execution completion or failure message of certain steps
- All these above-mentioned processes shall be developed although some may not be applicable in this provided dataset but could be for another dataset

2. Prediction

Build a prediction model for [SalePrice]

Requirement

- Use Python to develop the prediction model, choose your own features and algorithm to achieve optimal results
- Deliver your solution in (Jupyter) notebook or HTML format

3. Bonus Points

- If you deliver the solution (1) in an interactive and web-based interface that allows user to select items to be processed and able to view the processed results directly on web
- If you deliver the solution (2) in an interactive and web-based interface that allows user to adjust parameters or inputs so that the results can be fine-tuned or improved