Assignment: Data Quality Management

Due date : December 13, 2020 11:59pm

Deliverables:

1. A report (with introduction, body and conclusion) describing the issue, your analysis and your main findings in a pdf format (10 to 20 pages). The report can be generated with RMarkdown, Word, etc.

<div align="center">AND</div>

2. The original RMarkdown file.

Dataset:

The dataset contains retail data from 45 stores.

Store: the ID of the store.

Data: the weekly balance sheet date.

Holiday: a binary variable informing if the considered week is Holyday or not.

Type: the type of the store.

Size: the size of the store.

Dept: the department number of the store.

Sales: the volume of sales for the corresponding week.

Temperature: the average temperature in the region for the corresponding week.

Fuel_Price: the price of the fuel in the region for the corresponding week.

Promotion: price reduction for the corresponding store and week. The are 5 promotion categories.

CPI: the consumer price index of the corresponding region and week.

Unemployment: the unemployment rate in the region for the corresponding week.

Objective:

Your client is interested in explaining the variable **Sales** based on the given features. His question is: what are the determinants of performance of my stores?

Please mind that this is a chronological data (data that varies in time).

Some helpful packages:

tidyverse, reshape2, lubridate. Please feel free to use any packages you think helpful.