Project #2

**EDA/Data Cleaning Phase**

1. You will be using a data set accessed via the link [https://www.kaggle.com/competitions/store-sales-time-series-forecasting/data](https://nam12.safelinks.protection.outlook.com/?url=https%3A%2F%2Fwww.kaggle.com%2Fcompetitions%2Fstore-sales-time-series-forecasting%2Fdata&data=05%7C01%7Clkhinkis%40buffalo.edu%7C540a51fa42ac4506e5a308db3a05ade2%7C96464a8af8ed40b199e25f6b50a20250%7C0%7C0%7C638167568284852561%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C3000%7C%7C%7C&sdata=AjGjucDOuosoKeJfEAY3Zd4dSMpA3XztPNoJSqFqmJw%3D&reserved=0). The data is contained in the oil.csv file.
2. Plot the time series as is.
3. Read the literature and find out how to fill the missing data. Impute the data using your preferred method.
4. Plot the time series with imputed data. Do you see a trend and/or seasonality in the data?

**Modelling Phase**

1. Learn about the ETS models and about Holt-Winters models (provide all relevant specifics with respect to theoretical aspects and running them). This will expand your toolkit of the candidate models.
2. Based on your answer to the question 4, suggest suitable model(s) for the data.

**Evaluation Phase**

1. Run the models and check their adequacy.
2. Compare the models’ performance by the metrics that you think are relevant. Try to identify a model with a low RMSE.
3. The submission format is the same as the format for the homeworks. You need to submit two files, rmd file and a knit pdf file.

Due date 12/5.