Overview

Work on these in your groups

For all questions below, you may "trial and error" a number of approaches before settling on preferred model. I would like for you to **submit / present all major iterations of models that you try**, have a short discussion on their pros / cons and why you discarded some models in favor of others, rather than only submit / present the 'best model'.

One tip is to try to place all model building code in self-sufficient function calls. In this way, you can just rerun models with minor tweaks / different inputs whilst expending minimal (real world) time.

Another time saving step is to get your code to directly output "final presentation quality graphics" directly, and therefore zeroize the amount of time you spend formatting things.

Assignment should take ~1 hour if you are able to 'reuse' code from class, not more than 2 hours otherwise

1. **Question 1**: Use natural gas consumption data from https://www.eia.gov/dnav/ng/ng_cons_sum_dcu_nus_m.htm
    a. We will focus on "Volumes Delivered to Consumers" and its 5 components ("Residential", … "Electric Power")
    b. Using aggregate data series from 2001 to 2021, build and evaluate a time series model in this way:
        i. Perform time series decomposition
        ii. Use a benchmark model on seasonal component and another on trendcycle
        iii. Evaluate goodness of fit for model using techniques we learnt in class (e.g goodness of fit metrics, Ljung-Box tests, etc) .
    c. Perform same process as for (b)(i) and (ii), but on 5 individual components separately, before adding up forecasts from 5 individual models to get a total forecast for aggregate series.
        i. Hint: Get your code to print out numerical values for the forecasted series in an array or list. Then just add up those 5 arrays / lists to get an aggregate forecast. You can compare residuals, etc from here.
    d. Discuss comparative goodness of fit from (b) and also from (c).

2. **Question 2**: Download historical stock prices of the S&P 500 ETF ("SPY") for the last 20 years: https://finance.yahoo.com/quote/SPY/history?p=SPY
    a. Produce plots of data and form hypotheses on trends / cycles / seasonality as well as other artifacts
    b. Consider and implement (or discard) the relevance of data transformation techniques such as Box Cox, calendar etc corrections, modelling % changes instead of levels etc
    c. Split data into training and test set based on your discretion and guidelines we overviewed in class. Explain choices
    d. Use time series decomposition followed by various benchmark methods to forecast the training set and examine results on the test set. What method does best?
    e. Examine residuals of your preferred method and test them for time series information

3. **Question 3**: Textbook, Chapter 6.9 question 2 (on plastics data): https://otexts.com/fpp2/decomposition-exercises.html

4. **Question 4:** Textbook, Chapter 6.9 question 6 (on the bricks data). Same link as for Q3

Groups 1 to 4 will present during regular class time on Class 4.

Specifically, group 1 will present Q1, group 2 to present Q2, group 3 for Q3, and lastly group 4 presents Q4 in class.

For avoidance of doubt, here are revised and finalized groupings (reproduced):

DSA 301 2024 Groupings			
Group	Members	Group size	HW Presentation slot
1	Al Isthifah, Jeffrey Chan, Shaun Ng, Matthew Yap, Matthew Goh	5	1
2	Fang Qi, Xiao Hui, Matthew Cheng,, Iyer Jahnavim	4	1
3	You Yun, Qianyun Gao, Gloria Goh, Qistina Purnamasari	4	1
4	Nina Tan, Su Ying, Calista Lim, Nicholas How, Takyeon Kwon	5	1
5	Abigail, Danika Kaur, Xavier Boon, Julian Rong	4	2
6	Wei Qiang, Keith Yong, Yong Xiang, Samuel, Clarice Chan	5	2
7	Sharafinaz, Jin Rui, Chuan Wei, Caleb, Deviyani	5	2
8	Song Xi, Zihao, Wen Lin, Ee Ting, Thejas Puttur	5	2
9	Jester Teo, Tristan Julius, Swee Khai, Ian Yaoting, Teng Ling	5	2
10	Cassandra Chua, Syvester Song, Rayner Yeoh, Janias Tan	4	2