

Можно использовать любой софт с открытым кодом: python, r, julia, gretl, всё, что душе угодно из открытого софта. Работу можно выполнять командой из двух человек или в одиночку. Из трёх нельзя. Из четырёх тоже нельзя.

1. «Взять языка». Возьми месячный временной ряд. Можно не один, а несколько, чтобы использовать дополнительные ряды как предикторы основного. Не бери цены финансовых инструментов, так как их приращение плохо прогнозируется в силу эффективности рынка. Разумно взять реальные показатели. Если очень хочется работать с финансовыми данными, можно взять волатильность, она прогнозируется хорошо. Максимально чётко укажи, откуда взяты ряды. Если ряды парсились, то приведи код.
2. «Намалевич». Построй графики рядов, графики автокорреляционных функций, графики с нарезкой ряда на годы для иллюстрации сезонности.
3. «Твиттер». Кратко прокомментируй полученные графики. Являются ли ряды сезонными? есть ли тренд? стационарны ли ряды? есть ли точки излома? растёт ли амплитуда колебаний ряда? Возьми логарифм ряда, если душа тянется к логарифму.
4. «Двенадцать месяцев». Необходимо исследовать качество прогнозов минимум 6 моделей:
 - Наивная, $\hat{y}_{t+1} = y_t$.
 - Сезонная наивная, $\hat{y}_{t+1} = y_{t+1-12}$.
 - SARIMA(1, 1, 1)(1, 0, 0)[12].
 - Алгоритм Хиндмана-Хандакара автоматического подбора SARIMA.
 - ETS(AAA)
 - ETS с автоматическим выбором по AIC.

Сравни качество прогнозов по средней абсолютной ошибке MAE на тестовой выборке. В качестве тестовой выборки возьми последний год наблюдений, в качестве обучающей — все остальные наблюдения.

На бонусный балл: сравни качество прогнозов по средней абсолютной ошибке MAE на один шаг вперёд с помощью кросс-валидации. Кросс-валидацию проводи растущим окном, начав с окна в 80% исходной выборки.

Подсказка. В питоне могут помочь функции NaiveForecaster, ARIMA, AutoARIMA, ExponentialSmoothing и AutoETS из sktime. В r нужные модели есть в пакете fable, а графики поможет построить feasts. В julia ARIMA и ETS реализованы в пакете StateSpaceModels.

5. «Кто на свете всех милее?» Выбери наилучшую модель и построй график прогнозов для неё на один год вперёд используя все 100% наблюдений как обучающую выборку.

Ссылки:

- sktime: <https://www.sktime.org>.
- Прогнозы в sktime, в конце про кросс-валидацию: https://www.sktime.org/en/latest/examples/01_forecasting.html.
- Картинки для кросс-валидации: <https://www.kaggle.com/cworsnup/backtesting-cross-validation-for-time>
- Изложение алгоритма Хиндмана-Хандакара, да и вся книжка Forecasting Principles and Practice: <https://otexts.com/fpp3/arima-r.html>.

6. «Хожение в народ». Скачай панельные данные RLMS. Можешь выбрать данные по домохозяйствам, а можешь по индивидам, <https://www.hse.ru/rlms/>. Регистрация там бесплатная и без смс. Описание огромного количества переменных есть в <https://www.hse.ru/rlms/code>. К сожалению, данные RLMS выложены в закрытых форматах. Это маленький позор, я с ним пытался бороться, но пока безуспешно. Если уж никак не удаётся загрузить большую панельку из приватного формата сразу в открытый софт, открой её в родном закрытом софте и экспортируй нужные переменные в свободный .csv.
7. «Кому на Руси жить хорошо?». Сформулируй пусть не особо глубокий, но всё же вопрос. Уровня «Помогает ли потребление огурцов домохозяевами предсказать доход главы семейства?» вполне достаточно. В данных RLMS много пропусков, поэтому погоня за большим количеством предикторов приведёт к выбрасыванию всех наблюдений. Для данного игрового задания одного предиктора достаточно.
8. «Сделай красиво!» Визуализируй данные с целью графического ответа на поставленный вопрос.
9. «Три сестры». Оцени три модели: сквозную модель, FE-модель, RE-модель. С помощью подходящих тестов выбери наилучшую.

Подсказка. В питоне панельки оценивают с помощью <https://bashtage.github.io/linearmodels/>, в r — https://cran.r-project.org/web/packages/plm/vignettes/A_plmPackage.html, в julia — <https://nosferican.github.io/Econometrics.jl/stable/> или <https://github.com/FixedEffects/FixedEffectModels.jl>.
10. «Ответ на главный вопрос». Ответь на поставленный вопрос.

Наставления в добрый путь храброму падавану:

- Работу можно сдать в следующих видах. В виде читабельного отчёта в .pdf, приложив исходный .tex. В виде исполняемого .ipynb. В виде .html отчёта, сгенерированного из .Rmd или .jmd, приложив исходник.

Ссылки: с .ipynb большинство знакомо, поэтому <https://github.com/JunoLab/Weave.jl>, <https://bookdown.org/yihui/rmarkdown-cookbook/>.
- Приложи исходные данные, ряды и нужную для исследования часть панельки, в виде .csv. Не надо прикладывать всю скаченную панельку, но только лишь нужную часть.
- Мелкие детали, отсутствующие в условии, заполни самостоятельно, чётко описав свой выбор.
- Дедлайн: мягкий — 16 июня в 21:00. После мягкого — штраф минус 30%, жёсткий — 17 июня в 21:00, <https://classroom.github.com/a/MKA14xZ7>.
- Да пребудет с тобой Сила!