Non-Financial Time Series Project Guidelines

Objective

Students will select and analyze a real-world non-financial time series dataset by applying various time series modeling techniques. The goal is to perform exploratory analysis, apply appropriate models (e.g., ARMA, ARIMA, Seasonal ARIMA), evaluate model fit, and generate forecasts. The selected dataset should exhibit clear trends and/or seasonality.

Submission Requirements

Each group must submit:

- **Lightning Talk Presentation** (4 minutes):
  - Briefly describe the dataset and its relevance.
  - Justify the modeling approach and model selection criteria.
  - Present key findings and insights from the analysis.
  - Conclude with the implications of your results.
- **Jupyter Notebook**: A well-documented notebook containing:
  - All steps of the analysis, from data exploration to forecasting.
  - Code, explanations, and interpretation of results.

Tasks

1. Selecting a Non-Financial Time Series Dataset

Each group must select their own real-world non-financial time series dataset. The dataset must have a frequency of either hourly, daily, weekly, monthly, or quarterly observations. The length of the dataset should be adjusted accordingly:

- **Hourly**: At least 3 months of data.
- **Daily**: At least 1 year of data.
- **Weekly**: At least 3 years of data.
- **Monthly**: At least 5 years of data.
- **Quarterly**: At least 10 years of data.

The dataset must exhibit some trend and/or seasonality. Examples of suitable datasets include:

- Weather data (e.g., temperature, precipitation, air quality).
- Energy consumption or electricity demand.
- Website traffic or social media engagement.
- Public health data (e.g., hospital admissions, disease incidence).
- Transportation data (e.g., traffic flow, public transit usage).

Students can find suitable datasets from the following sources (but you can use others too):

- `https://datahub.io/` (Various open data repositories)
- `https://www.data.gov/` (US government datasets, including environmental and public health data)
- `https://www.ncdc.noaa.gov/cdo-web/` (NOAA climate and weather data)
- `https://ourworldindata.org/` (Long-term public health, environmental, and demographic data)
- `https://data.europa.eu/` (European Union open data portal)

- `https://www.ons.gov.uk/` (UK Office for National Statistics)
- `https://ec.europa.eu/eurostat` (Eurostat datasets on energy, transport, and environment)
- `https://datos.gob.es/en` (Spanish government open data portal)
- `https://www.comunidad.madrid/gobierno/datos-abiertos` (Comunidad de Madrid open data portal)
- **R Packages**: `UKgrid` (UK electricity demand data), `gtrendsR` (Google Trends time series data)

Groups must provide a brief description of their dataset, its source, and its relevance.

2. Exploratory Data Analysis (EDA)

Perform an initial analysis of the time series:

- Plot the time series and discuss any visible patterns or trends.
- Check for seasonality and periodicity.
- Compute summary statistics and visualize autocorrelation (ACF, PACF).
- Handle missing values and potential outliers.

3. Model the Time Series Using Various Approaches

Apply multiple models to the data and compare their performance:

- Fit different models (e.g., ARMA, ARIMA, Seasonal ARIMA).
- Justify your choice of model parameters (e.g., order selection via AIC/BIC).
- Compare in-sample fit and diagnostics (e.g., residual analysis).

4. Model Selection and Diagnostics

Evaluate models based on:

- Goodness-of-fit criteria (AIC, BIC, RMSE, MAPE, etc.).
- Residual diagnostics (ACF/PACF of residuals, normality checks).
- Overfitting concerns and interpretability of results.

5. Generate Forecasts and Interpret Results

- Provide in-sample and out-of-sample forecasts.
- Assess forecast accuracy using appropriate error metrics.
- Discuss the practical implications of your forecasts.

Additional Notes

- Clearly label and annotate all plots.
- Write concise explanations for every step in the Jupyter Notebook.
- Ensure the presentation is engaging and to the point.