# Time Series Analysis Course Project

Prof. Dr. Md Rezaul Karim
Department of Statistics and Data Science
*Jahangirnagar University*

Fall-2025

## 1 Overview

This project requires you to apply time series analysis techniques to real-world economic and business datasets. You will analyze the provided data, test for stationarity, build forecasting models (ARIMA, SARIMA, and Exponential Smoothing), and compare their performance.

## 2 Dataset Description

You are provided with a dataset (`timeseries_project_data.csv`) containing 60 months of observations (January 2019–December 2023) for four different time series:

1. **Retail_Sales**: Monthly retail sales in million USD

2. **Energy_Consumption**: Monthly energy consumption in GWh (Gigawatt-hours)

3. **Stock_Price_Index**: A stock market index value

4. **Unemployment_Rate**: Monthly unemployment rate as a percentage

## 3 Project Tasks

### 3.1 Part 1: Exploratory Data Analysis (20 points)

Perform exploratory data analysis for **all four** time series:

1. Plot each time series and describe the patterns you observe (trend, seasonality, cyclical behavior, irregularities).

2. Calculate and report descriptive statistics (mean, standard deviation, minimum, maximum) for each series.

3. Create ACF and PACF plots for each series up to 24 lags.

4. Decompose each series into trend, seasonal, and residual components. Plot the decomposition and comment on each component.

5. Based on your visual analysis, which series appear stationary? Which series show clear seasonal patterns? Explain your reasoning.

## 3.2   Part 2: Unit Root Testing (15 points)

Test each time series for stationarity:

1. Conduct the Augmented Dickey-Fuller (ADF) test on all four series. Report the test statistic, $p$-value, and your conclusion at the 5% significance level.

2. Conduct the KPSS test on all four series. Report the test statistic, $p$-value, and your conclusion at the 5% significance level.

3. Create a summary table showing all test results.

4. For any non-stationary series, apply first differencing and re-test. Report whether the differenced series is stationary.

5. Do the ADF and KPSS tests agree for all series? If not, discuss possible reasons for the disagreement.

## 3.3   Part 3: ARIMA Modeling (25 points)

Choose **TWO** series **randomly** from the dataset and build ARIMA models:

1. For each chosen series, determine the appropriate order of differencing ($d$) needed.

2. Using ACF and PACF plots of the differenced series, suggest appropriate values for $p$ and $q$.

3. Fit at least **three different** ARIMA models with different $(p, d, q)$ combinations for each series.

4. Compare the models using AIC and BIC. Create a table showing the parameter values and information criteria for all fitted models.

5. Select the best model for each series and explain why you chose it.

6. For your best models, perform residual diagnostics:

   - Plot the residuals over time
   - Create ACF plot of residuals
   - Conduct the Ljung-Box test on residuals (report test statistic and $p$-value)
   - Create a Q-Q plot and conduct a normality test

7. Do the residuals behave like white noise? Explain.

8. Generate 12-month ahead forecasts with 95% confidence intervals and plot them along with the original data.

## 3.4   Part 4: SARIMA Modeling (20 points)

Choose the series that shows the **strongest seasonal pattern**:

1. Justify your choice of series based on Part 1 analysis.

2. Determine the seasonal period $S$ from the data.

3. Examine the ACF and PACF at seasonal lags. What values of $P$ and $Q$ do they suggest?

4. Fit at least **two different** SARIMA models with different parameter combinations.

5. Compare your SARIMA models with the best non-seasonal ARIMA model (from Part 3 if you used the same series, or fit a new ARIMA model). Create a comparison table showing AIC, BIC, and RMSE.

6. Which model performs better? Why do you think the SARIMA model is superior (or not)?

7. Perform residual diagnostics on your best SARIMA model.

8. Generate 12-month ahead forecasts and compare them visually with ARIMA forecasts for the same series.

### 3.5    Part 5: Exponential Smoothing Models (20 points)

Use the same series as Part 4 (the one with strong seasonality):

1. Fit the following models:

   - Simple Exponential Smoothing
   - Holt's Linear Trend method
   - Holt-Winters Additive Seasonal method
   - Holt-Winters Multiplicative Seasonal method

2. Calculate in-sample fit measures (MSE, MAE, MAPE) for each model.

3. Create a comparison table showing AIC and fit measures for all four ES models.

4. Which ES model performs best? Why?

5. Generate 12-month ahead forecasts using your best ES model.

6. Compare the ES forecast with the SARIMA forecast from Part 4. Which one would you trust more and why?

7. Which model provides the most accurate forecasts? Does this match your expectations ?

## 4    Submission Requirements

You must submit:

### 4.1    Written Report (80% of grade)

A professionally formatted PDF report (maximum 25 pages) containing:

- **Title page**: Project title, your name, student ID, date

- **Executive summary** (1 page): Key findings and recommendations

- **Analysis sections**: Complete answers to all tasks in Parts 1–5, including:
  - All required plots and tables with proper captions
  - Detailed interpretation of results
  - Clear explanations of your decisions and conclusions

- **Conclusions** (1–2 pages): Summary of findings, best models, limitations, and recommendations

- **References**: Any sources consulted

## 4.2   Code Files (20% of grade)

Submit your analysis code as:

- Jupyter notebook (.ipynb) OR Python script (.py) OR R script (.R) OR Eviews Output File

- Code must be well-commented and organized

- Code must be reproducible (someone else should be able to run it and get the same results)

- Include comments explaining each major step

# 5   Grading Rubric

Table 1: Grading Breakdown

| Component | Points | Criteria |
|---|---|---|
| Part 1: EDA | 20 | Completeness, quality of plots, depth of insights |
| Part 2: Unit Root Tests | 15 | Correct application and interpretation |
| Part 3: ARIMA | 25 | Model selection, diagnostics, forecasting |
| Part 4: SARIMA | 20 | Seasonal identification, comparison, evaluation |
| Part 5: ES Models | 20 | Multiple models, comparison, forecasting |
| *Analysis Total* | *100* | |
| Report Quality | 40 | Clarity, organization, professional presentation |
| Code Quality | 10 | Organization, comments, reproducibility |
| **Total** | **150** | *Scaled to 100* |

# 6   Evaluation Criteria

Your work will be evaluated based on:

- **Correctness**: Are the methods applied correctly?

- **Completeness**: Are all tasks addressed thoroughly?

- **Interpretation**: Do you explain what the results mean, not just report numbers?

- **Critical thinking**: Do you compare models thoughtfully and justify your choices?

- **Presentation**: Is your report clear, well-organized, and professional?

- **Code quality**: Is your code readable, organized, and reproducible?

# 7   Important Notes

- You are free to use Python, R, or any statistical software for your analysis

- Focus on **interpretation** – don't just report numbers, explain what they mean

- Use visualizations effectively to support your conclusions

- Compare models carefully and provide reasoned justifications for your choices

- Start early – time series analysis requires experimentation

## 8   Important Dates

- **Submission Deadline**: December 25, 2025, 11:59 PM

- **Late submissions**: 20% penalty per day

## 9   Academic Integrity

This is an **individual project**. You may discuss general concepts with classmates, but all analysis, code, and written work must be your own. Copying from others or from online sources without attribution constitutes plagiarism and will result in a zero grade.

*Good luck with your analysis!*