DEGREE: MSc Data Analytics
Module: Fundamentals of Data Analytics

Assignment Title: Time Series Forecasting: A Practical Approach to Data-Driven Decision Making

Learning Meet: Set exercise (not under exam conditions)

Word Limit: 3000 words (+/- 300)

Weighting: 40%

Issue Date: 5/9/2025

Submission Date: 29/9/2025

Feedback Date: 17/10/2025

Plagiarism:

When submitting work for assessment, students should be aware of the InterActive/Canvas guidance and regulations concerning plagiarism. All submissions should be your own, original work.

You must submit an electronic copy of your work. Your submission will be electronically checked.

Learner declaration
I certify that the work submitted for this assignment is my own and research sources are fully acknowledged. **Student signature:** **Date:**

Harvard Referencing:

The Harvard Referencing System must be used. Wikipedia, UKEssays.com or similar websites must **not** be used or referenced in your work.

Introduction

Learning Outcomes:

LO1. Demonstrate an understanding of basics of linear algebra, calculus, statistics and algorithms to structure basic solutions of business using data.

LO2. Use of statistics for business decision making and the ways to source relevant data to assist decision making & problem solving.

LO3. Solve problems using derivatives, transcendental functions, and integration. Integrate the concepts of foundations of data analytics for problem solving.

This assignment focuses on applying time series regression techniques to predict future values of a continuous target variable.

Dataset Description:

- Data Type: Time series
- Target Variable: Continuous (e.g., energy consumption in kWh, temperature in °C)
- Size: Over 1000 data points per dataset
- Features: Includes timestamp, temperature, humidity, wind speed, and other environmental or economic indicators

Justification for Selection:

- Real-world relevance (e.g., energy forecasting, climate modeling)
- Sufficient data volume and feature richness
- Publicly available and well-documented

Task Phases:

Task 1: Exploratory Data Analysis (EDA):

- Summary statistics (mean, median, standard deviation)
- Time series plots to visualize trends and seasonality
- Correlation heatmaps to identify relationships between variables
- Seasonal decomposition to detect patterns

Task 2: Data Preprocessing and Feature Engineering

Objective: Prepare the data for regression modeling and create relevant features.
Steps Taken:

- Missing Values: Handled using forward fill and interpolation
- Outliers: Detected using IQR and z-score methods
- Time-Based Features Created:
    - Day of week
    - Month
    - Season
    - Lag features (e.g., previous day's consumption)
- Correlation Analysis: Pearson correlation matrix
- Normalization: StandardScaler used for models sensitive to scale

Task 3: Regression Model Development and Evaluation

Objective: Build and compare at least two time series regression models.
Instructions:

Models Implemented:

1. Linear Regression
2. Random Forest Regressor

Data Splitting:

- Training Set: 70%
- Validation Set: 15%
- Test Set: 15%

Evaluation Metrics:

- Mean Squared Error (MSE)
- Mean Absolute Error (MAE)
- R-squared (R^2)

Feature Importance:

- Random Forest used to rank features
- Lag features and temperature were most influential

Hyperparameter Tuning:

- GridSearchCV used on validation set

Final Results:

- Random Forest outperformed Linear Regression in all metrics
- Predictions on test set showed good alignment with actual values

Task 4: Conclusions and Business Recommendations

Conclusions:

- Clear seasonal and weekly patterns in energy consumption
- Temperature and previous consumption are strong predictors

Strategic Recommendations:

1. Optimize energy distribution during peak seasons using forecasts
2. Implement dynamic pricing based on predicted demand
3. Use predictive maintenance to reduce downtime and improve efficiency

Submission Instructions:

- The dataset is provided for you.
- Ensure that your report is clear, well-organized, and visually appealing. It should explain the steps you took for data cleaning. Include a detailed introduction, methodology, findings, and conclusions in your report.
- Prepare your document using the BSBI assignment template available on Canvas.
- Remember to cite all sources using the Harvard referencing style and strictly adhere to academic integrity guidelines.
- Submit your assignment electronically by the specified deadline.
- Include an accessible **link** to your Colab notebook implementation in the assignment, ensuring that your code is well-commented and properly documented. Your Google Colab link should be accessible or share via GitHub.
- Refer to the Essay Guide available on Canvas for further instructions.