COMPETENCIES

4030.7.2 : **Time Series Analysis**
The graduate applies time series models in generating forecasts.

INTRODUCTION

As a data analyst, you will assess continuing data sources for their relevance to specific research questions throughout your career.

In your previous coursework, you performed data cleaning and exploratory data analysis on your data. You have seen basic trends and will now build more sophisticated statistical models.

For this task, you will select one of the data sets and associated data dictionaries from "Data Files and Associated Dictionary Files" in the Web Links section.

After you choose your organizational data set, you will review the data dictionary and considerations related to the raw data file you have chosen and prepare the data for time series modeling. You will then analyze that data set using time series modeling, create visualizations, generate forecasts, and deliver the results of your analysis.

Note: If you have trouble accessing the link, copy and paste the link directly into your web browser.

REQUIREMENTS

Your submission must be your original work. No more than a combined total of 30% of the submission and no more than a 10% match to any one individual source can be directly quoted or closely paraphrased from sources, even if cited correctly. The originality report that is provided when you submit your task can be used as a guide.

You must use the rubric to direct the creation of your submission because it provides detailed criteria that will be used to evaluate your work. Each requirement below may be evaluated by more than one rubric aspect. The rubric aspect titles may contain hyperlinks to relevant portions of the course.

*Tasks may **not** be submitted as cloud links, such as links to Google Docs, Google Slides, OneDrive, etc., unless specified in the task requirements. All other submissions must be file types that are uploaded and submitted as attachments (e.g., .docx, .pdf, .ppt).*

Choose one of the data files from the "Data Files and Associated Dictionary Files" web link below and use the information to complete the following:

Part I: Research Question

A. Describe the purpose of this data analysis by doing the following:

1. Summarize **one** research question that is relevant to a real-world organizational situation captured in the selected data set and that you will answer using time series modeling techniques.
2. Define the objectives or goals of the data analysis. Ensure your objectives or goals are reasonable within the scope of the scenario and are represented in the available data.

Part II: Method Justification

B. Summarize the assumptions of a time series model including stationarity and autocorrelated data.

Part III: Data Preparation

C. Summarize the data cleaning process by doing the following:
   1. Provide a line graph visualizing the realization of the time series.
   2. Describe the time step formatting of the realization, including *any* gaps in measurement and the length of the sequence.
   3. Evaluate the stationarity of the time series.
   4. Explain the steps you used to prepare the data for analysis, including the training and test set split.
   5. Provide a copy of the cleaned data set.

Part IV: Model Identification and Analysis

D. Analyze the time series data set by doing the following:
   1. Report the annotated findings with visualizations of your data analysis, including the following elements:
      - the presence or lack of a seasonal component
      - trends
      - the autocorrelation function
      - the spectral density
      - the decomposed time series
      - confirmation of the lack of trends in the residuals of the decomposed series
   2. Identify an autoregressive integrated moving average (ARIMA) model that accounts for the observed trend and seasonality of the time series data.
   3. Perform a forecast using the derived ARIMA model identified in part D2.
   4. Provide the output and calculations of the analysis you performed.
   5. Provide the code used to support the implementation of the time series model.

Part V: Data Summary and Implications

E. Summarize your findings and assumptions by doing the following:
   1. Discuss the results of your data analysis, including the following points:
      - the selection of an ARIMA model
      - the prediction interval of the forecast
      - a justification of the forecast length
      - the model evaluation procedure and error metric
   2. Provide an annotated visualization of the forecast of the final model compared to the test set.
   3. Recommend a course of action based on your results.

Part VI: Reporting

F. With the information from part E, create your report using an industry-relevant interactive development environment (e.g., an R Markdown document, a Jupyter Notebook). Include a

PDF or HTML document of your executed notebook presentation.

G. Cite the web sources you used to acquire third-party code to support the application.

H. Acknowledge sources, using in-text citations and references, for content that is quoted, paraphrased, or summarized.

I. Demonstrate professional communication in the content and presentation of your submission.