

NLM3 – NLM3 TASK 1: TIME SERIES MODELING

ADVANCED DATA ANALYTICS – D213

PRFA – NLM3

TASK OVERVIEW

SUBMISSIONS

EVALUATION REPORT

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.

File Restrictions

File name may contain only letters, numbers, spaces, and these symbols: ! - _ . * ' ()

File size limit: 200 MB

File types allowed: doc, docx, rtf, xls,xlsx, ppt, ptx, odt, pdf, csv, txt, qt, mov, mpg, avi, mp3, wav, mp4, wma, flv, asf, mpeg, wmv, m4v, svg, tif, tiff, jpeg, jpg, gif, png, zip, rar, tar, 7z

RUBRIC

A1:RESEARCH QUESTION

NOT EVIDENT

The submission does not provide a summary of 1 research question.

APPROACHING COMPETENCE

The submission summarizes 1 research question, but the research question is not relevant to a realistic organizational situation or cannot be addressed using the selected data set and time series modeling techniques.

COMPETENT

The submission summarizes 1 research question, and the research question is relevant to a realistic organizational situation and can be addressed using the selected data set and time series modeling techniques.

A2:OBJECTIVES OR GOALS

NOT EVIDENT

The submission does not define the objectives or goals of the data analysis.

APPROACHING COMPETENCE

The submission defines the objectives or goals of the data analysis, but 1 or more of the objectives or goals are not clear or reasonable within the scope of

COMPETENT

The submission clearly defines *each* of the objectives or goals of the data analysis. *Each* objective or goal is reasonable within the scope of the scenario and represented in the available data.

the scenario or are not represented in the available data.

B: SUMMARY OF ASSUMPTIONS

NOT EVIDENT

The submission does not summarize the assumptions of a time series model.

APPROACHING COMPETENCE

The submission summarizes the assumptions of a time series model, but the summary does not include stationarity or auto-correlated data. Or 1 or more of the assumptions contain inaccuracies.

COMPETENT

The submission accurately summarizes *each* of the assumptions of a time series model. The summary includes stationarity and autocorrelated data.

C1: LINE GRAPH VISUALIZATION

NOT EVIDENT

The submission does not provide a line graph.

APPROACHING COMPETENCE

The submission provides a line graph, but the graph does not align with the chosen data set. Or the line graph is not clearly labeled or is an incomplete realization of the time series.

COMPETENT

The submission provides a line graph that aligns with the chosen data set. The line graph is clearly labeled and is a complete realization of the time series.

C2: TIME STEP FORMATTING

NOT EVIDENT

The submission does not describe the time step formatting of the realization.

APPROACHING COMPETENCE

The submission inaccurately describes the time step formatting of the realization, or the description does not include *any* gaps in measurement or the length of the sequence.

COMPETENT

The submission accurately describes the time step formatting of the realization, including *any* gaps in measurement and the length of the sequence.

C3: STATIONARITY

NOT EVIDENT

The submission does not evaluate the stationarity of the time series.

APPROACHING COMPETENCE

The submission evaluates the stationarity of the time series, but the evaluation does not align with the chosen data set or the

COMPETENT

The submission accurately evaluates the stationarity of the time series, and the evaluation aligns with the chosen data set and the research question.

research question. Or the evaluation contains inaccuracies.

C4: STEPS TO PREPARE THE DATA

NOT EVIDENT

The submission does not explain the steps used to prepare the data for analysis.

APPROACHING COMPETENCE

The submission explains the steps used to prepare the data for analysis, but the steps are incomplete or do not include the training or test set split. Or 1 or more of the steps are not related to preparing for time series modeling.

COMPETENT

The submission explains the steps used to prepare the data for analysis. *Each* step is complete and includes the training and test set split and relates to preparing for time series modeling.

C5: PREPARED DATA SET

NOT EVIDENT

The submission does not provide a copy of the cleaned data set.

APPROACHING COMPETENCE

The submission provides a copy of a cleaned data set, but the data set is not fully prepared.

COMPETENT

The submission provides a copy of a fully prepared, cleaned data set.

D1: REPORT FINDINGS AND VISUALIZATIONS

NOT EVIDENT

The submission does not report the annotated findings with visualizations of the data analysis.

APPROACHING COMPETENCE

The submission reports the annotated findings with visualizations of the data analysis, but the findings are incomplete or inaccurate. Or the report does not include 1 or more of the listed elements.

COMPETENT

The submission reports the annotated findings with visualizations of the data analysis. The findings are complete and accurate and include *each* of the 6 listed elements.

D2: ARIMA MODEL

NOT EVIDENT

The submission does not identify an ARIMA model.

APPROACHING COMPETENCE

The submission identifies an ARIMA model, but the model does not account for the ob-

COMPETENT

The submission identifies an ARIMA model that accounts for the observed trend and seasonality of the time series data. The

served trend or seasonality of the time series data, or the identified model is not appropriate for the data set or chosen research question.

identified ARIMA model is appropriate for the data set and chosen research question.

D3:FORECASTING USING ARIMA MODEL

NOT EVIDENT

The submission does not perform a forecast using the derived ARIMA model from part D2.

APPROACHING COMPETENCE

The submission provides a forecast using the derived ARIMA model, but the forecast is inaccurate. Or the model is not the one identified in part D2.

COMPETENT

The submission provides an accurate forecast using the derived ARIMA model identified in part D2.

D4:OUTPUT AND CALCULATIONS

NOT EVIDENT

The submission does not provide the output or calculations of the analysis performed.

APPROACHING COMPETENCE

The submission provides the output and calculations of the analysis performed, but the output or calculations are incomplete or contain inaccuracies.

COMPETENT

The submission provides *all* the complete and accurate output and calculations of the analysis performed.

D5:CODE

NOT EVIDENT

The submission does not provide the code used to support the implementation of the time series model.

APPROACHING COMPETENCE

The submission provides the code used to support the implementation of the time series model, but the code is incomplete, or there are errors when the code is run.

COMPETENT

The submission provides the code used to support the implementation of the time series model. The code is complete and runs without errors.

E1:RESULTS

NOT EVIDENT

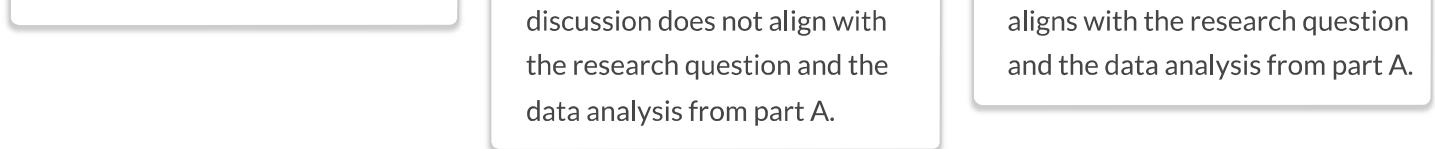
The submission does not discuss the results of the data analysis.

APPROACHING COMPETENCE

The submission inaccurately discusses the results of the data analysis or does not include 1 or more of the given points. Or the

COMPETENT

The submission accurately discusses the results of the data analysis and includes *each* of the 4 given points. The discussion



E2: ANNOTATED VISUALIZATION

NOT EVIDENT

The submission does not provide an annotated visualization of the forecast of the final model compared to the test set.

discussion does not align with the research question and the data analysis from part A.

aligns with the research question and the data analysis from part A.

APPROACHING COMPETENCE

The submission provides an incomplete annotated visualization of the forecast of the final model compared to the test set, or the forecast does not align with part D1.

COMPETENT

The submission provides a complete annotated visualization of the forecast of the final model compared to the test set, and the forecast aligns with part D1.

E3: RECOMMENDATION

NOT EVIDENT

The submission does not recommend a course of action based on the results.

APPROACHING COMPETENCE

The submission recommends a course of action based on the results, but the recommendation is not appropriate based on the research question or the results of the data.

COMPETENT

The submission recommends an appropriate course of action based on the results as they relate to the research question.

F: REPORTING

NOT EVIDENT

A report is not created. Or the PDF or HTML document of the executed notebook presentation is not provided.

APPROACHING COMPETENCE

The report is inaccurately created in an industry-relevant interactive development environment. Or the PDF or HTML document of the executed notebook presentation is incomplete or does not align with the data analysis of the report.

COMPETENT

The report is accurately created in an industry-relevant interactive development environment. The PDF or HTML document of the executed notebook presentation is provided, is complete, and aligns with the data analysis of the report.

G: SOURCES FOR THIRD-PARTY CODE

NOT EVIDENT

The submission does not list *any* web sources.

APPROACHING COMPETENCE

The submission cites only *some* of the web sources used to ac-

COMPETENT

The submission cites *all* web sources used to acquire third-party code, and *all* web sources

H:SOURCES

NOT EVIDENT

The submission does not include both in-text citations and a reference list for sources that are quoted, paraphrased, or summarized.

quire third-party code. Or 1 or more of the web sources are not reliable.

are reliable. And the citation information includes the author, date, title, and source location, as available.

I:PROFESSIONAL COMMUNICATION

NOT EVIDENT

Content is unstructured, is disjointed, or contains pervasive errors in mechanics, usage, or grammar. Vocabulary or tone is unprofessional or distracts from the topic.

APPROACHING COMPETENCE

Content is poorly organized, is difficult to follow, or contains errors in mechanics, usage, or grammar that cause confusion. Terminology is misused or ineffective.

COMPETENT

Content reflects attention to detail, is organized, and focuses on the main ideas as prescribed in the task or chosen by the candidate. Terminology is pertinent, is used correctly, and effectively conveys the intended meaning. Mechanics, usage, and grammar promote accurate interpretation and understanding.

WEB LINKS

Data Files and Associated Dictionary Files

If you have trouble with the link, copy and paste the URL directly into your web browser.