

# RFN1 — RFN1 TASK 3: TIME SERIES MODELING

MACHINE LEARNING — D603

PRFA — RFN1

Preparation

**Task Overview**

Submissions

Evaluation Report

## COMPETENCIES

### 4163.1.3: Applies Time Series Models

The learner applies time series models in generating forecasts.

## INTRODUCTION

In your previous coursework, you have performed data cleaning and exploratory data analysis on your data. You have seen basic trends and patterns and now want to start building more sophisticated statistical models. In this task, you will use time series modeling. You will explore time series models and their implications.

For this task, you will select one of the data dictionary and dataset files to use for your report from the following options:

churn\_clean.csv and Churn Data Considerations and Dictionary.pdf

medical\_clean.csv and Medical Data Considerations and Dictionary.pdf

After you choose your organizational dataset, 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 dataset using time series modeling, create visualizations, generate forecasts, and deliver the results of your analysis.

## REQUIREMENTS

Your submission must represent your original work and understanding of the course material. Most performance assessment submissions are automatically scanned through the WGU similarity checker. Students are strongly encouraged to wait for the similarity report to generate after uploading their work and then review it to ensure Academic Authenticity guidelines are met before submitting the file for evaluation. See [Understanding Similarity Reports](#) for more information.

### Grammarly Note:

Professional Communication will be automatically assessed through Grammarly for Education in most performance assessments before a student submits work for evaluation. Students are strongly encouraged to review the Grammarly for Education feedback prior to submitting work for evaluation, as the overall submission will not pass without this aspect passing. See [Use Grammarly for Education Effectively](#) for more information.

### Microsoft Files Note:

Write your paper in Microsoft Word (.doc or .docx) unless another Microsoft product, or pdf, is specified in the task directions. Tasks may not be submitted as cloud links, such as links to Google Docs, Google Slides, OneDrive, etc. All supporting documentation, such as screenshots and proof of experience, should be collected in a pdf file and submitted separately from the main file. For more information, please see [Computer System and Technology Requirements](#).

*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.*

A. Create your subgroup and project in GitLab using the provided web link by doing the following:

- Clone the project to the IDE.
- Commit with a message and push when you complete each requirement listed in part D.

*Note: You may commit and push whenever you want to back up your changes, even if a requirement is not yet complete.*

- Submit a copy of the GitLab repository URL in the "Comments to Evaluator" section when you submit this assessment.
- Submit a copy of the repository branch history retrieved from your repository, which must include the commit messages and dates.

B. 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 dataset and that you will answer using time series modeling techniques.
2. Define the objectives or goals of the data analysis. Ensure that your objectives or goals are reasonable within the scope of the scenario and are represented in the available data.

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

D. 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 used to prepare the data for analysis, including the training and test set split.
5. Provide a copy of the cleaned dataset.

E. Analyze the time series dataset by doing the following:

1. Report the annotated findings with visualizations of your data analysis, including the following elements:
  - 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 E2.

4. Provide the output and calculations of the analysis you performed.

F. Summarize your findings and assumptions by doing the following:

1. Discuss the results of your data analysis, including the following:
  - 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 that includes the following:
  - the original output with the new prediction line and confidence cone
  - correct labeling
3. Recommend a course of action based on your results.

G. 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.

H. Record the web sources used to acquire data or segments of third-party code to support the analysis. Ensure the web sources are reliable.

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

J. 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, pptx, 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

### A:GITLAB REPOSITORY

#### NOT EVIDENT

The submission does not provide a GitLab repository.

#### APPROACHING COMPETENCE

The submission provides a subgroup and project created in GitLab, but 1 or more of the given actions are not completed, or they are completed incorrectly.

#### COMPETENT

The submission provides a subgroup and project correctly created in GitLab, and *all* of the given actions are completed correctly.

### B1:RESEARCH QUESTION