

# FINAL PROJECT GUIDELINES

**Purpose:** The purpose of this assignment is for you to experience the complete life cycle of a project based on time series data analysis. You will utilize the data analytic strategies that you might use later in an analyst or business role. These strategies will be applied to a forecasting problem that you are interested in answering. The forecasting steps require you to build a forecast from the best fit model and assess the accuracy of that forecast. For more information on the forecasting project cycle, consult ‘the five basic steps in a forecasting task’ outlined in section 1.6 (page 21-22) of your textbook.

**Skill:** The purpose of this assignment is to help you practice the following skills that are essential to your success in this course and in professional life beyond school:

- \* Formulate a problem statement related to a forecasting question.
- \* Design a strategy to answer the forecasting question.
- \* Collect, clean, and use publicly available time-series data.
- \* Evaluate various forecasting models to get the best possible forecast.
- \* Select the best model based on model selection criteria and accuracy measures.
- \* Execute appropriate statistical tests to obtain better forecasting accuracy.
- \* Interpret, communicate, and present forecast results in a report.

**Knowledge:** This assignment will also help you to become familiar with the following valuable content knowledge in this discipline.

- \* Defining a forecasting problem.
- \* Utilizing a statistical program to compute, visualize, and analyze time-series data in economics, business, and the social sciences.
- \* Performing exploratory analysis.
- \* Selecting an appropriate statistical model among alternative models.
- \* Validating the selected statistical model.
- \* Interpreting models using parameters.
- \* Forecasting based on the selected statistical model.
- \* Assessing the accuracy of forecasts.
- \* Interpreting and communicate results effectively.

**Task:** The final project will consist of a brief report between 10-12 double-spaced pages, including relevant tables and figures. To begin, you may choose your time-series dataset. Choose a dataset that has a time component and a variable to analyze over time. Choose a dataset that you would like to analyze according to the techniques that are outlined in the textbook and were discussed in class. Find the model that you believe fits your data best and build a forecast from this model. Assess the validity of that forecast. You must follow ‘the five basic steps in a forecasting task’ outlined in section 1.6 of your textbook.

Your report should be structured as follows:

1. Abstract (no more than 250 words): A summary of your basic findings.
2. Introduction (1-3 pages): A brief introduction/motivation to the problem at hand, relevant details about the data, additional relevant scientific information from searching the web, for example, and what is to be addressed.
3. Data Description (at most one page): A brief introduction to the data, data sources, and variable definitions.
4. Statistical Methods (1-2 pages): A discussion and justification of the methods you have used to analyze the data, and how you went about analyzing the data. Don't forget to describe in some detail how and why the particular model was selected.
5. Results (2-3 pages): A presentation of the results of your analysis. Interpretations should include a discussion of statistical versus the practical importance of the results.
6. Discussion (1-2 pages): A synopsis of your findings and any limitations your study may suffer from. Present conclusions in terms that non-statisticians will understand. Quantitative and qualitative aspects should be discussed.

*Your report should be brief and to the point! It should be written in a language that is understandable to the scientific community.*

#### **Criteria for Success:**

1. Explanation: Excellent reports will use a data set a student is interested in and apply domain knowledge that describes why the forecasting is useful or important. This importance will be explained clearly in the introduction and conclusion of the report and will be evident to the reader. Reports should be clear and concise and use scientific language.
2. Process-oriented: Good reports will also follow the 5 basic steps in a forecasting task described in the textbook. Be sure to apply all the tools used and models discussed this semester to determine the best forecast, and clearly show why this forecast is the best using the accuracy measures. Reports should follow the task steps described above in the correct order.
3. Evidence-based: Use accuracy measures to back up the selected model. Show why alternative models are not as efficient with your specific dataset.

# Forecasting Final Project Rubric

Category	Excellent	Good	Satisfactory	Needs Improvement
<b>Interpretation</b> <i>Ability to explain information presented in mathematical forms (e.g., equations, graphs, diagrams, tables, words) in analytics language</i>	Effectively explains models that were selected and <b>why</b> they were selected using analytics language	Attempts to explain models, does not fully explain <b>why</b> they were selected using some analytics language	Does not fully explain what models were selected or <b>why</b> they were selected. Occasional analytics language is used.	Does not state why models were selected, does not state <b>why</b> they were selected. No analytics language is used.
<b>Representation</b> <i>Ability to convert relevant information into various mathematical forms (e.g., equations, graphs, diagrams, tables, words)</i>	Uses figures and equations effectively to enhance discussion	Uses figures and equations to add to discussion but could be more effective	Attempts to use figures and equations but does not use them effectively	Does not use figures or equations in discussion.
<b>Calculation</b> <i>Ability to correctly deduce most accurate model and use accuracy measures</i>	Correctly and comprehensively uses models and accuracy measures Calculations are also presented elegantly (clearly, concisely, etc.)	Models and accuracy measures attempted are essentially all successful and sufficiently comprehensive to solve the problem.	Models and accuracy measures attempted are either unsuccessful or represent only a portion of the calculations required to comprehensively solve the problem	Models and accuracy measures are attempted but are both unsuccessful and are not comprehensive.
<b>Application / Analysis</b> <i>Ability to make judgments and draw appropriate conclusions based on the quantitative analysis of data, while recognizing the limits of this analysis</i>	Analysis is thorough and complete, and coherent throughout the discussion	Analysis is mostly complete, does not capture all information relevant to discussion	Analysis is incomplete, and does not effectively or coherently draw conclusions	No analysis or analytics thinking is present.
<b>Assumptions</b> <i>Ability to make and evaluate important assumptions in estimation, modeling, and data analysis</i>	Explicitly describes assumptions of the unique dataset and models and provides compelling rationale for why each assumption is appropriate. Shows awareness that confidence in final conclusions is limited by the accuracy of the assumptions.	Explicitly describes assumptions and provides compelling rationale for why assumptions are appropriate for this dataset and models.	Explicitly describes assumptions for this dataset and models chosen.	Attempts to describe assumptions for this dataset and models chosen.
<b>Communication/Presentation</b> <i>Expressing quantitative evidence in support of the argument or purpose of the work (in terms of what evidence is used and how it is formatted, presented, and contextualized)</i>	Uses quantitative information in connection with the argument or purpose of the work, presents it in an effective format, and explicates it with consistently high quality. Document is well organized, error free, and neat.	Uses quantitative information in connection with the argument or purpose of the work, though data may be presented in a less than completely effective format or some parts of the explication may be uneven.	Uses quantitative information but does not effectively connect it to the argument or purpose of the work. Document could be organized better	Does not correctly communicate quantitative information or does not build a cleanly presented final document.