

STAT GR5221/GU4221 Spring 2023 Project

1 Choose a Topic

Write-up a short paragraph to describe your project and your objective(s). **This description is due on Canvas by April 13th. The project is due during the reading week.**

Students should submit:

- i. Describe your topic and objective(s)
- ii. Describe your collected data (if applicable)
- iii. Are you working in a group or individual?
 - Full list of group partners and UNIs

2 Potential Project Topics

Students should choose a project that suits their personal interests and directly utilizes topics presented in Time Series Analysis (4221/5221). Potential project topics:

- I. Predictive Modeling - Univariate Time Series (CH1,CH2, CH3, CH5, CH6)
 - ARMA
 - ARIMA
 - SARIMA
 - Self study on seasonality
- II. Predictive Modeling - Multivariate Time Series (CH8)
- III. Spectral Approach to Time Series (CH4)
- IV. Predictive Modeling - Financial time series (CH7)
 - ARCH
 - GARCH
 - ARMA-ARCH
 - ARMA-GARCH
- V. Predictive Modeling - Machine Learning and Time Series (Similar to STAT4241/5241):
 - Long Short Term Memory
 - Hidden Markov Models
- VI. State-Space Models (CH9)
 - Kalman Filter

- State-Space Models with Missing Observations (EM algorithm)

VII. Continuous time models (CH11)

- CAR
- CARMA

VIII. Theoretical Topic

- Asymptotics
- Any topic covered in class
- See Brockwell and Davis, Theory and Methods Text
- See Hamilton Text
- Simulation Study

IX. Stationarity Tests (DF, KPSS, ...)

X. Functional Time Series Analysis Data (Kokoska Text)

XI. A combination of the above ideas

XII. Many other choices..

Please contact Prof. Young for recommendations and guidance on any project type.

3 Individual and Group Option

Students can work in groups with up to 3 students, i.e., individual project, group of 2, group of 3. For students who choose to work in groups, they are required to fill out an assessment form to describe each members' contribution.

4 Expectations

One objective of the project is to provide students with a *data analysis component* to the GU4221/GR5221 course. The GU4221/GR5221 project is not intended to heavily discriminate final grades, in fact, a large majority of the class is expected to earn full points on this assignment. The exams are designed to discriminate final grades. Students should produce some final product that can be discussed in job interviews, or similar.

5 Rubric for Predictive Modeling

Please note: If a students do not choose the predictive modeling option, then they do not have to follow the below rubric exactly. Try to complete your write-up based on your best judgment.

This project is graded with a high level of tolerance. If a student gives a reasonable effort, they should get close to full credit. If a student gives a large effort, they should get full credit. If a student gives little effort, it is still possible for them to earn some of the points based on the project rubric.

	Points	Earned points
Exploratory analysis	10	
Time series model and modeling process	30	
Model validation	10	
Forecasting	10	
Quality of the report	20	
Attached code file	10	
Other	10	
Total	100	

5.1 Exploratory Analysis [10 points]

Students should provide some exploratory data analysis in their project.

5.2 Time Series Model and Modeling Process [30 points]

Students should clearly describe their chosen model and how they arrived at their final model. This section is arguably the most important component of the predictive modeling choice.

5.3 Model Validation [10 points]

Students must assess the performance of their statistical model, (or models) by running a traditional model validation procedure. A common data split is recommended for the predictive model and/or metrics such as AIC, BIC, etc.

5.4 Forecasting [10 points]

Include a section on forecasting.

5.5 Quality of the Report [20 points]

The final write-up should be concise and well organized. It's important for students to be rewarded for the quality of their write-up and how well they conveyed information to the graders/TAs. If a student produces very high quality work, they should earn full credit in this section. If a student gives an unorganized mess that is very hard to grade, then they will earn low marks on this section. A short write-up is not a bad thing if the information is presented well and a long overly detailed report is not ideal.

5.6 Attached Code File [10 points]

Students should prepare an organized **R script** (or **Rmd**) file that complements the written report (or **Python** file). **Do not** copy and paste the code into your appendix. Only include very important code, or no code, in your final report. Please upload the script or similar on Canvas by the due date. Graders and TAs will double check that students provided a code file to compliment their final report. This could be in any programming language.

5.7 Other [10 points]

This section allows the grader/TA to give their subjective opinion on a student's project. If a student satisfied all project components but their overall project is not great, the grader can give low or medium marks in this section. If the student did an outstanding job, then give them full marks in this section.

6 Websites for Finding Data

- [Kaggle](#)
- [UCI Machine Learning Repository](#)
- [NYC Open Data](#)
- [London DATASTORE](#)
- Yahoo Finance
- Google Finance
- [NCAR Data Archive](#).
- [Search on your own..](#)