**DSCI 5340 - Predictive Analytics and Business Forecasting**

**Project**

Due: June 30, 2021

***Group Project***

There will be a project that will require team work. Related handouts will be distributed in class and related datasets will be posted on Canvas. You will be asked to form teams of 2 members. You will have to select a business problem that interests you (the instructor will suggest a problem in case you run out of ideas.) Problem data/facts should be real, obtained from verified sources, and correspond to realistic situations. Your team will prepare a PowerPoint presentation, to be presented in class. Your team may also submit a written section that includes your SAS code or any coding for other analytical methodology used in this project. The presentation should be no more than 30 slides.

The final project is a chance to reflect the knowledge gained in this course through applying the techniques to practical scenarios. It will utilize **time-series analysis** to construct a forecast model using SAS. You are expected to collect real-world data, and have it approved before you begin your analysis. The purpose is ultimately to make a predictive model that accurately forecasts a given variable from other values.

*Sell – Tell –Compel Method*

You should use the following structure to present the findings from your projects. This is the most common method for business/analytics presentations. It will help you organize your project in an easily presentable manner.

***Sell* your audience on the importance of your research**.

The first section is an introduction to the problem that you will be solving. Provide background and context for your presentation. This is also where you will establish the theory for the correlation of variables in your model. Include in this section approx. 5 academic and/or journalistic publications that outline both the need for this predictive ability and why you decided to use the method that you did. (E.g. This article says that heart disease is preventable yet kills thousands every year; these articles say that these factors influence heart disease; therefore, controlling these factors that correlate with heart disease will help prevent it)

***Tell* the story of your scientific process/research**.

The next section includes the steps that you took to reach your conclusion. Start by describing your data, then identify the variables that you used. Then state the hypothesis (H0 and HA). Explain the model that you are using and how you came to this. Show not only your final model, but each model that you tried and why you included, transformed or excluded variables (multicollinearity, norms violated, high p-values, etc.). Please validate your model with residual analysis and/or collinearity (VIF) table. Finally, show the model statistics in table form and any graphs of your model. This section is the heart of your presentation and should represent the majority of your slides.

***Compel* your audience to action**.

The final section is the presentation of your conclusion and implications for the future. If your null hypothesis is rejected, explain what impact this has on the subject of your research. If the null hypothesis is not rejected, then suggest variables to add or remove from your model or how would you modify your theory in future research.

***Topics to consider:***

* Health and healthcare
* Humanitarian operations
* Disaster relief
* Education
* Social services
* Environment
* Sustainability
* Sharing economy
* Transportation
* Urban planning
* Fraud, collusion, and corruption
* Government policy
* Poverty
* Privacy
* Cyber security
* Crime and terrorism

***Recommended Journals to use for forecasting:***

* International Journal of Forecasting
* Forecasting (MDPI)
* Journal of Forecasting

**NOTE:** You are not limited to this list, but you are required to check if the journals and papers you choose are ranked well in their respective fields. Above are some forecasting journals, however depending on the topic you choose, there are journals available within that topic related to forecasting and regression. For example, if you choose “Health and healthcare”, you will find forecasting related problems within medical journals too.

**Grading Rubric**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Exceeds Standards | Meets Standards | Approaching standards | Does not meet standards |
| Introduction  (7 pts) | * Thorough review of existing research * Exhaustive theory and hypothesis * Justified methodology | * Gives context appropriate to research * Provide existing knowledge, theory and proposed methodology | * Missing some aspect of intro; purpose, theory, methodology | * Does not provide any context for research |
| Literature Review  (3 pts) | * > 5 peer-reviewed academic articles relevant to topic /variables | * 3-5 peer-reviewed articles and 2-3 journalist articles | * 1-3 peer-reviewed articles and 2-3 journalist articles   relevant to topic /variables | * <1 peer-reviewed articles and <2 journalist articles * Not relevant to topic /variables |
| Hypothesis statement  (3 pts) | * Statement matches theory * Includes all variables and relationships to be tested | * Clearly defined and relevant null and alternative hypotheses | * Statement does not match theory * Null and alternative are incongruent | * No hypothesis provided |
| Methodology  (25 pts) | * Constructs consistent and supported by theory * Includes multiple iterations, including non-significant findings * Utilizes sophisticated modeling (ARIMA, interactions, etc.) | * Appropriate to hypothesis * Includes appropriate variables/ relationships * Provides critical values, p-values, F-score or other meaningful measures | * Model does not match data/hypothesis * Incorrect coefficients * Model/ residuals not tested for multicollinearity, normality | * Grossly inadequate model * Missing key parameters such as R-squared, ACF, PACF, etc. |
| Code  (15 pts) | * Provide all code (preferably in SAS) * Utilize secondary analysis (residuals) * Utilizes visualization, metadata procedures | * Provide coding used for data analysis * If not using SAS provide screenshots or other documentation | * Missing some lines of code * Code is inaccurate * Does not use SAS for any portion | * Code is not included * Missing significant portions |
| Conclusion  (7 pts) | * Synthesizes results and theory; explain the relationship between variables * Relates model to hypotheses, matches variables * Explains implication of model/ results | * Summarizes results * Refers to theory * Provides implication for future actions or studies | * Incorrect conclusion * Does not refer to hypothesis or theory * Implications do not represent results | * Does not provide results * Does not relate research model to practical implementation or topic |

Please remember that this is a graduate-level research project. The quality of writing, analysis and presentation should be at the level of academic journals. In addition, plagiarism or any other actions that violate the academic integrity pledge will not be tolerated. You have resources available to you, including the campus library and the TA, so please take advantage of them.

I look forward to seeing what you can produce and best regards.

**Mr.  Kiran Kumar Reddy Chadagonda**

[KiranKumarReddyChadagonda@my.unt.edu](mailto:KiranKumarReddyChadagonda@my.unt.edu)

**Tuesdays from 12:00 to 2:00 pm and Friday from 12:00 to 2:00 pm**.

**Meeting ID: 935 080 0297**