

ITEC 621 Predictive Analytics

**Project Instructions**

(in Teams of 3 to 4)

*Prof. Espinosa – Last updated 8/7/2024*

**Table of Contents**

<b>Background .....</b>	<b>1</b>
<b>CRISP-DM Overview .....</b>	<b>2</b>
<b>Data .....</b>	<b>3</b>
<b>Requirements .....</b>	<b>4</b>
<b>Modeling Method, Model Specification and Procedures .....</b>	<b>6</b>
<b>Project Deliverables .....</b>	<b>7</b>
Project Template.....	7
Deliverable 1 (10 pts): Project Proposal.....	7
Deliverable 2 (10 pts): Preliminary Data Analysis Report.....	8
Deliverable 3 (5 pts, individually graded): Meet with Professor .....	10
Deliverable 4 (65 pts): Final Report .....	10
Deliverable 5 (10 pts): Class Presentation .....	13

**Background**

Business analytics is not just about analyzing data. Rather, it is about solving and then communicating solutions for business problems using credible data as your evidence. This requires teamwork, critical analysis, and a compelling upfront articulation of the specific business problem or analytics question you are solving as well as a clear and concise report of the findings and recommendations. Accordingly, the main goal of this project is to prepare for your practicum projects by giving you an opportunity to put into practice what you have learned in class.

The predictive analytics project will be done in teams of approximately four students. We expect that all team members will contribute equally and that everyone will take the opportunity to learn from each other.

**Important:** students **must** work only with **models** and **R coding methods** covered in class. If your team wishes to explore other methods or use R coding methods not covered in the class materials, you need to describe your methods to me and get my approval.

## CRISP-DM Overview

The **Cross Industry Standard Process for Data Mining (CRISP-DM)** is the most widely used networks for data mining and analytics projects. I provide an overview of this network as an informational piece. The CRISP-DM maps closely to the **INFORMS' Job Task Analysis (JTA)** framework (<http://info.informs.org/analytics-body-of-knowledge>; [Amazon](#)), which is also followed widely in management science projects. The deliverables for this project are largely based on a shorter version of the CRISP-DM framework, so this will provide you with a good reference frame. In essence, the CRISP-DM framework and the corresponding JTA activities and domains include the following activities:

- **Business Understanding (CRISP-DM 1)**
  - (*JTA Domain I*) Formulate the **business question** to be answered or problem to be solved. All business analytics projects must be driven by business needs or business value propositions. This requires clearly stating the business question or problem you hope to solve upfront.
  - (*JTA Domain II*) Translate the business question into the respective **analytics question**. Not all business questions or problems are amenable to analytics solutions. The project report must specify how or why analytics is the appropriate approach to address the business question or problem.
- **Data Understanding (CRISP-DM 2)**
  - Data acquisition and pre-processing can take as much as 80% of the analytics project effort.
  - (*JTA Domain III*) Acquire and identify relationships in the data. This step involves **acquiring the data** (e.g., ETL or Extract-Translate-Load) and then doing a substantial amount of **descriptive analytics**, including things like (as appropriate): descriptive statistics, correlation analysis, ANOVA, distribution curves, visual plots and other graphs, and other related analysis (e.g., cluster analysis). NOTE: Predictive analytic modeling should not start until you have developed a thorough understanding of the data. In fact, this phase may uncover issues and relationships in the data that you did not anticipate, thus leading to reformulation of the analytics question.
- **Data Preparation (CRISP-DM 3)**
  - (*JTA Domain III*) Harmonize, re-scale and clean data, as needed. Data sets often need to be split, merged, sub-sampled (for large data sets), and cleansed. This step involves all **data pre-processing** activities, such as: re-structuring the data (e.g., normalizing scales, centering, aggregating, etc.); addressing issues of missing data; and acquiring and merging other related data.

- **Modeling (CRISP-DM 4)**
  - Select the appropriate analysis *methodology* and tools, exploring various model *specifications*, and then building the respective models. In this course we use R as the primary analytical tool.
  - (*JTA Domain IV*) Methodology Selection. Most of the course is focused on *method selection* (e.g., OLS regression, Logistic regression, Ridge or LASSO, trees, etc.). Candidate models should be identified based on the analytics goals: interpretation, inference and/or prediction). For this project, students need to focus on models that are relatively interpretable and then select the model that has better predictive accuracy, based on cross validation test error or deviance.
  - (*JTA Domain V*) Model Building. Another area of focus in this course is on *model specification*, e.g., linear, polynomial, interactions, or variable selection. The initial set of predictors to be used in the model must be driven by business domain knowledge. But then this set should be narrowed down or refined using statistical methods like cross-validation testing.
- **Evaluation (CRISP-DM 5)**
  - Note that this phase is not about evaluating the models. This happens in the Modeling phase above. **This phase is about evaluating the extent to which the analysis has answered the business and analytics questions framed in phase 1.** For this project, we will focus on the following:
  - Interpretation of Results: the final project reports must provide a focused *interpretation of results*, in terms of effects observed, fit statistics, and predictive power of the final model.
  - An important part of this interpretation is providing a well-documented answer to the business and analytics question - how will these findings be applied or drive strategy?
  - It is also important that you tell a compelling story in your report. *Storytelling* is one of the most important skills in business analytics. Remember, this is not a statistics class, but a business class. You must tell a compelling story for your audience. The story must be backed up by your findings.
- **Deployment (CRISP-DM 6)**
  - (*JTA Domain VI*) For this project, deployment will focus on turning in your *written report*, with the necessary interpretation and stories articulated in step 5 above.

**Important note: not all projects lead to amazing findings!!** A model that shows no effects can offer very interesting insights. It all depends on how you rationalize the lack of effects from a business point of view. Along the same lines, this project is not so much about what you analyzed and found, but about how effectively you described to your readers the motivation for your study, your method evaluation and selection process and what the implications of your findings are from a business perspective.

## Data

Any dataset **not used** previously for class **lectures, exercises, homework or term projects** can be used for this semester's project. Students are expected to identify an interesting external data set

to work with. In the past, many students have used Kaggle data sets used in competitions, but there are many sources of public data. Proprietary data sets can only be used with permission of the owner of the data set. It is fine to use data from your practicums, if you have it, and use this project as an opportunity to work with your client's data. Unless the data is proprietary, teams must submit the actual datasets with their final projects so that the professor can replicate some of your work when grading. If you use a dataset that you used previously in another course, you need to provide the name of the course and instructor. You also need to briefly describe what you did in the earlier project and your analysis for the current project must be substantially different than what you did in that course.

## Requirements

**Teams** of no more than **4** students **must** evaluate:

- **3** different **modeling methods** (e.g., OLS, Ridge, Logistic, LDA, trees, etc.)
  - Start with OLS for quantitative outcomes or Logistic regression for classification outcomes, which would be your first model.
- **2** different **model specifications** for **each** method (e.g., different predictor subsets; polynomial, log or other transformations; interactions, etc.).
  - Start with a specification that includes the outcome variable in your analytics question and predictors selected based on good business rationale, e.g.,  $y \sim x_1 + x_1 + x_3$
- Test the assumptions of your initial OLS or Logistic model and based on the results, select:
  - 2 additional models to evaluate.
  - Ideally, the second model method could be a parametric model that addresses the issues you encountered in your assumption testing (e.g., Ridge, LASSO, PCR)
  - The third model method could be a non-parametric model with good interpretability (e.g., Random Forest Tree, Boosted Tree).
  - Analyze a second specification based on your assumption tests and intuition, for example:
    - ✓ A larger or smaller set of predictors
    - ✓ A model with transformations (e.g., a log model, a polynomial model, etc.)

There are 3 key points to keep in mind:

**POINT #1:** The 2 model specifications selected above should be used in **each** of the 3 modeling methods above. The best approach is to fit the first model using OLS for a quantitative prediction, or Logistic regression for a classification prediction, using both model specifications. Then, depending on your results and assumption testing, fit the same 2 specifications using two other models.

**POINT #2:** All team members must contribute their fair share of the analysis. The expectation is that each member will take the lead on one particular modeling method or transformation. Each team will provide feedback on their teammates during the semester to evaluate how each member contributed to the project. I will ask students to report on their individual contribution to the team's project.

**POINT #3:** While you will be evaluating and testing 6 different models (3 model methods x 2 specifications), you should **only report on the final model methods and specification selected**. If you selected one model for interpretability and a different model for predictive accuracy, then you need to report on both. You need to discuss your model selection process, including any fit statistics and cross-validation test results for all 6 models and discuss how you arrived at your final selection. Once you select a particular modeling method and specification, you need to fit your final model with the **full dataset**. If you wish to include output from alternative models and specifications, you can do that in an appendix.

**POINT #4:** You don't need to include all your R code and output. There is simply no time to review all your R code and output. But you can provide some of the key output in appendices, if you wish.

**POINT #5 – IMPORTANT:** All your deliverables must contain subsections matching the items requested below. Please don't submit a monolithic report with no section titles. It only makes it more difficult to evaluate your project, so help me give you the credit you deserve for your good work, rather than guessing.

## Modeling Method, Model Specification and Procedures

Analytics Question	Model Types	Modeling Methods (Evaluate 3)	Model Specification (Evaluate 2 for each model method)	Testing, Procedures and Comments
Quantitative	Parametric	<ul style="list-style-type: none"><li>Regression: OLS, Ridge, LASSO, PCR, PLS</li></ul>	<ul style="list-style-type: none"><li>Predictors (feature) selected based on business knowledge.</li><li>Predictors (features) selected by statistical analysis (e.g., variable selection, ANOVA, stepwise, etc.)</li><li>Variable transformations (e.g., polynomials, log models, lagged or time-series models, piecewise models, splines, ranks, Box-Cox transformations, etc.)</li><li>Categorical to binary variable transformations are not different specifications but required transformations.</li><li>Centering and standardizing are not model specifications, but different interpretations of results.</li></ul>	<ul style="list-style-type: none"><li>WLS is not really a different method, but a correction for OLS’ heteroskedasticity.</li><li>GLM is not really a different method, but an estimation method applicable to several methods.</li><li>Step variable selection (Forward, Backward and Stepwise) are not model methods, but evaluation procedures for variable selection.</li><li>Bootstrap, RSCV, LOOCV and KFCV are not model methods, but cross-validation tests applicable to any method.</li><li>You can use <math>R^2</math>, Adjusted <math>R^2</math>, model p-value, coefficient p-values, Cp, 2LL, AIC, BIC, etc. to analyze the statistical fit. But all candidate models and specifications <b>must be compared with a cross-validation testing method</b>.</li><li>Changing tuning parameters don’t yield different models or specifications but help select the optimal tuning of models (e.g., tree size, Ridge shrinkage, PCR components).</li></ul>
	Non-Parametric	<ul style="list-style-type: none"><li>Regression trees, bootstrap aggregation trees, random forest trees, boosted trees, neural networks, k-nearest neighbors</li></ul>		
Classification	Parametric	<ul style="list-style-type: none"><li>Binomial and multinomial Logistic regression</li><li>LDA and QDA</li></ul>		
	Non-Parametric	<ul style="list-style-type: none"><li>Classification trees, bootstrap aggregation trees, random forest trees, boosted trees, support vector machines, neural networks</li></ul>		

**Note:** only models and specifications covered in this course are allowed for the course project. Any team contemplating other methods not covered in the course must obtain permission from the instructor.

# Project Deliverables

## Project Template

This project must be completed using the project template provided on Canvas. Use the same template and project document for all deliverables. The idea is to complete your project incrementally in a single document. This template contains all the required sections for the project. And departures from the template must be discussed with and approved by me.

This project has **5 deliverables**:

**Deliverable 1 (10 pts): Project Proposal** (1 page, single-spaced or 2 pages double-spaced)

Deliverable 1 Objectives:

- Prepare and articulate your business and analytics questions and business case.
- Identify one or more potential data sets to use for the project.

Details:

A project proposal is due around the mid-semester point, per the class schedule. The goal in this first deliverable is to get you started on your project early and provide the direction you are planning to take in your project. It is also an opportunity to get feedback on your project ideas.

**Previously used data:** If you use a dataset that you used previously in another course, you need to get approval from me. You also need to provide the name of the course where you used the data and the instructor's name. For approval, you also need to briefly describe what you did in the earlier project in your proposal and how your analysis for the current project will differ from the past project.

The proposal should contain the following sections:

- (1) **The business case** should lead to the business question. It should contain a well-articulated statement of the problem you are aiming to solve and a concise but well-supported rationale about the importance of this problem to your client and audience. That is, what is the value proposition of your project? What facts, data, or statistics support your case? This should not be a general statement glorifying your project, but one that can convince your audience that it is worth pursuing. The business case should lead naturally to the business question that follows.
- (2) **Business and analytics questions** – these are tightly related and you need to articulate both:
  - a) The **business question** is the general question that your managerial audience is seeking answers for and, therefore what drives your project. The business question needs to be articulated in plain English and somewhat general terms (e.g., which are the main factors leading to coronary heart disease? How can we reduce the spread of a pandemic?)
  - b) The analytics question is a more technical, specific and detailed version of your business question and should be stated in plain English. All predictive models follow the specification  $y \sim x_1 + x_2 + \text{etc.}$  The analytics question should be the plain English articulation of your focal model specification. That is, you must articulate the outcome (Y)

of your model and whether it is quantitative or classification, along with the focal predictors (X's) you plan to evaluate with your model. You don't need to mention all possible predictors in your data set, but just the most relevant predictors to your business question. Refrain from using cryptic variable names and use plain English terms instead.

- (3) **One or more possible datasets** identified for the project. The more specific the datasets you are contemplating the better. You need to specify the possible sources for your dataset (e.g., Kaggle, Google Datasets, etc.)

**Note:** if your team wishes to change the proposal after it has been submitted, it is OK, but you will need to prepare and submit a new proposal and any other deliverables you may have already submitted. You can include the revised proposal in the Deliverable 2 document below.

**Deliverable 2 (10 pts): Preliminary Data Analysis Report** (2 pages of text **MAX**, single-spaced, or 4 pages double-spaced, plus appendices with R output as needed)

Deliverable 2 Objectives:

- Revise and refine your business case and business and analytics questions.
- Decide on a final data set for the project and provide a brief description.
- Conduct basic descriptive analytics to familiarize yourself with the data and develop some understanding of its properties (e.g., means, variances, correlations, etc.).
- Conduct a preliminary analysis with OLS (for a quantitative outcome) or Logistic (for a classification outcome) regression to develop a basic understanding of the relationship between predictor and outcome variables.
- Conduct basic OLS or Logistic regression assumption testing to better understand your modeling options going forward.

Details:

This deliverable is intended to get you started early on your project model method and specification exploration. It is also meant to get you familiarized with the project data. You should think of this deliverable as an early draft of your final report. It is also one last opportunity to get feedback on the direction of your project.

**IMPORTANT:** your main text should contain only narratives. Place all statistical output and plots in appendices. All appendices must be appropriately referenced in the main text. **Only include appendices and R output if they support your main text narrative. Unrelated appendices and output are confusing and unnecessary.**

Before you do your preliminary data analysis, **revise and refine your project proposal as needed.** More specifically, refine your business case, business question and analytics question. Your Deliverable 2 report must contain these revisions. Everything in your proposal (i.e., business case, business question, analytics question, etc.) should also be part of your deliverable 2 report. That is, further develop your project document and don't write a new document from scratch.

In addition to your project revisions indicated above, your Deliverable 2 documents must also contain:



(3) **Data Set Description.** In Deliverable 1 you discussed possible data sets to use. For this deliverable, you need to have decided on the specific data set you will use in your project. You don't need to provide a full description of the dataset yet, but you need to provide enough information for your professor to understand what you are analyzing. There is no need to provide extensive data descriptions in this deliverable, but you need to indicate the data source, the main variables of interest to your study (using descriptive names rather than cryptic variable names), the total number of observations. Describe the respective variable type, unit of measurement, and a brief description for each variable of interest in your dataset. No need to describe all variables in the dataset, just the ones you plan to use. If your outcome is a binomial classification variable, indicate whether the data set is relatively balanced (i.e., with approximately equal number of positives and negatives) or unbalanced.

(4) **Descriptive analytics.** You must provide a brief discussion of the respective:

4.1 Descriptive statistics of key variables (i.e., means and standard deviations or variances)

4.2 Distribution of the outcome variable (e.g., normal, skewed, Poisson, binary, etc.), complemented by a histogram, if needed. There is no need to analyze the distribution of all predictor variables, unless you plan to transform some of them.

4.3 Correlation and co-variation analysis (e.g., correlation analysis, ANOVA, chi-square tests of independence), which can be complemented by plots and graphs, as needed.

4.4 Data preprocessing and transformations you are contemplating based on your descriptive analytics above.

**Important:** You should not only describe what your analysis above, but it is also important to briefly discuss why you did that analysis and what you learned from it to inform your upcoming predictive modeling (e.g., to do an initial inspection for possible multicollinearity, we computed the correlation matrix of the focal predictors and we found some high correlation values, raising multicollinearity concerns). That is, descriptive analytics should not be done in the vacuum, but should inform your modeling decisions. If a particular type of descriptive analytics does not inform your modeling decisions it does not need to be reported in this section.

(5) **Modeling Methods and Specifications.**

**5.1 Initial Model Specification.** Briefly articulate the outcome variables and predictors from your data set that you plan to use in your initial model specification. Indicate whether your outcome variable is quantitative, binary, or categorical. Note that this initial specification must be consistent with your analytics question above. In addition, you also need to discuss any variable transformations you are contemplating based on your discussion in 4.4 above. Your initial model specification must be based on business domain rationale, so please include a brief articulation of your rationale.

**5.2 Initial OLS or Logit Model Results.** Because all model explorations begin with either an OLS regression (for quantitative predictions) or a Logistic regression (for classification predictions), the preliminary data analysis report will include either an **OLS regression** if your outcome variable is quantitative, or **Logistic regression** if your outcome variable is classification. In either case you must include the predictors identified above. Later in the

project, you will refine this initial set of predictors through variable selection, best subsets, or other methods.

**5.3 Assumption Tests.** **Inspect** residual and other regression **plots**, as appropriate. If your model is quantitative, conduct the necessary assumption **tests** to evaluate adherence to the OLS regression **assumptions**, (e.g., multicollinearity, serial correlation if there is time data, heteroscedasticity, linearity). If your model is for classification, you only need to test for multicollinearity and serial correlation if there is time data.

Finally, provide a brief statement of your conclusions. Your discussion must articulate the types of possible models you will evaluate next and what you learned in general.

### **Deliverable 3** (5 pts, individually graded): **Meet with Professor**

Deliverable 3 Objectives:

- All team members must be in the meeting
- Evaluate teamwork and team member contributions to the project, which are individually graded.
- Discuss your project and any issues you may have encountered Prepare and articulate your business and analytics questions and business case to help you succeed in the project.

Details:

The full team needs to meet with me shortly after the Deliverable 2 submission. This meeting is **required**. Team members will be graded individually based on their understanding of the project issues, general contributions to the project, and participation in the Deliverable 3 meeting. In addition to these individual points, the final project grade may be adjusted upwards or downwards based on team evaluations and my evaluation of your contributions towards the project. ALL team members must attend this meeting, so please schedule your meeting with me ahead of time. A team member that does not attend this meeting will not earn the assigned points above. The purpose of this meeting is for me to learn how your team is working out, what is the general contribution of each team member, discuss any methodological issues or challenges you may have encountered, and for me to give you feedback on your Deliverable 2 and your project as a whole.

### **Deliverable 4** (65 pts): **Final Report** (4 to 5 pages of text, single-spaced, plus appendices with R output as needed)

Deliverable 4 Objectives:

- Final reporting

Details:

**IMPORTANT:** as should be clear by now, one important learning objective in the MS Analytics program is to be able to interpret analytics results and articulate them clearly to a business audience. This is called “**storytelling**”. In a recent documentary, I heard the expression “data has no emotion, stories do”. Without a compelling story, data analytics fall flat. It all boils down to

writing concisely to the point and clearly what your results mean for your client's business. "Storytelling" means that you are persuasively communicating information to address a problem the client faces and could be "a story" about agile solutions, concern for brand reputation and efficiency, gaining ground over the competition, or otherwise improving the business practices of a client to their advantage. In today's business environment, your role is critical, because all business "stories" must be backed up by empirical evidence from data, such as your interpretations of statistical output. Avoid grandiose statements and fluff. Get to the point right away because the paper space is limited and business people like succinct but informational writing. Without a story, simply presenting data will not appeal to emotions and will not drive effective business decision-making.

The final project report will be submitted as an analytics report prepared in MS Word or knitted with R Markdown as a Word or PDF document. Most of these sections should be an extension of your Proposal and Preliminary Data Analysis Report above. The final project report will contain the following sections:

- (1) **Business Case (5 pts.)**, as described above. Please review and update this section from prior deliverables, as needed.
- (2) **Business and Analytics Questions (5 pts.)**, as described above. Please review and update this section from prior deliverables, as needed.
- (3) **A description of the dataset (5 pts.)**, as discussed above. Please review and update this section from prior deliverables, as needed. *As indicated above, describe your* data source(s), the main variables of interest to your study contained in the data set, the total number of observations and, for each variable please describe its respective variable type, unit of measurement, and a brief description. Provide a brief description of your data set to help your audience understand it. In the narrative focus on key study variables, but also include and refer to a table in the appendix with all the variables used in the models. Please note that your dataset may contain hundreds of variables, but you only need to describe variables that are relevant to your analysis.
- (4) **Descriptive Analytics (10 pts.)**, as discussed above. Please review and update this section from prior deliverables, as needed. It is also important to provide a brief concluding paragraph at the end of this section articulating important aspects uncovered by your descriptive analytics of the data (i.e., visual plots, descriptive statistics, correlations, etc.). Rather than making general comments about the analysis, focus on the aspects that informed the development of your predictive models.
- (5) **Modeling Methods and Model Specifications you evaluated and selected (15 pts.)**, as described above. Please review and update sections 5.1, 5.2 and 5.3 from prior deliverables, as needed. In addition, complete the other sections:
  - 5.4 Model Candidates and Rationale.** Besides OLS or Logistic regression, provide a brief discussion of your model candidates. One of your models will be either OLS or Logistic regression, which you don't need to rationalize. But provide a brief rationale for considering any other models you analyzed.
  - 5.5 Model Specification Candidates and Rationale:** based on your variable selection and data pre-processing results above, provide a brief description and discussion of the model

specifications you are considering. You only need two model specifications (or more for larger teams as indicated above). The two models can be a small and a large model, a linear and a more complex transformed model, etc. The important thing is to rationalize why you selected these specifications. I recommended selecting the small or linear model specification with variables informed by the business domain, and the larger or more complex model informed by some variable selection or cross-validation analysis. It is important to not just describe these alternative model specifications, but also provide a brief rationale for why you considered these specifications.

**5.6 Cross-Validation Testing and Final Model Selection:** provide a brief description of the cross-validation test results from all the candidate models and specifications you considered and indicate your final model selection and the criteria you based your decision on. Important note: if your analytics goal is interpretability, your final model must be interpretable; if your goal is predictive accuracy, your final model has to have the lowest CV test error; if you have both goals, you can either select the most interpretable model with acceptable CV test error, or you could select two models, one to use for interpretation and one for predictive accuracy.

- (6) **Analysis of Results (10 pts).** Your analysis and results need to contain some narrative to allow your audience to understand what you did. A simple output, diagram and data dump with no explanation will receive very little credit. Every procedure, output and diagram need to be briefly but appropriately introduced first and briefly commented on its meaning after. Don't leave it up to the reader to interpret what you did. Also, vague and general discussions of results will receive little credit. Your narrative of results should be factual and specific, so it needs to be backed up by fit statistics, coefficient values and significance, etc. Most importantly, your analysis must answer the analytics question you articulated above. That is, you start your report by formulating a question you intend to answer, so therefore you must end by answering that question. Important: you don't need to describe and report on everything you did. Focus on the key aspects of your analysis and the results that contribute to your final interpretations and story. Generally speaking, you only need to discuss results in depth for your final model selection. If your final selection is two models, one for interpretation and another for accuracy, discuss the relevant aspects of each. For the remaining candidate models, focus on reporting cross-validation results and your rationale for not selecting them.
- (7) **Conclusions and Lessons Learned (5 pts.).** Business analytics is about gaining insights from business data for decision making. This is the section to articulate what insights you gained from your analysis. These **conclusions** must contain a discussion of:
- 7.1 **Conclusions from the Analysis.** These conclusions **must answer** the business question(s) stated in 1 above. Please be brief but concise and discuss the main insights obtained from your analysis.
- 7.2 **Project Issues, Challenges and Lessons Learned.** This final item is not as important for your project, but it is important for me and your peers to learn about your experience with this project. Please provide a brief statement of the main issues and challenges you faced in this project and what you learned from it, including things like data issues,

methodological challenges, do's and don'ts, and what you learned from this experience. You don't need to address all of this or write extensively. Just a few thoughtful and interesting statements would be sufficient.

**(8) Writing Quality, Formatting and Presentation (10 pts.).** We have heard over and over from recruiters, practicum managers and our ITEC advisory board that storytelling is everything. Managers are less interested in p-values and other fit statistics, and more interested in the meaning and insights you gained from your analysis, and how these influence business decisions and outcomes. Analytics projects, no matter how good, are not useful unless the analytics report is well-written and clearly articulated. Statistical output without sound commentary about the results and implications for business is of little use to decision-makers. Consequently, the insights from your story, along with the businesslike document formatting, presentation, and writing clarity of the report, free of grammatical errors and typos will be heavily considered as a critical part of your grade. Just as important, the entire report needs to flow, persuade, and be understandable to your audience.

### **Deliverable 5 (10 pts): Class Presentation** (5 to 6 slides of content)

Deliverable 1 Objectives:

- Prepare and deliver short presentations to share your work with your professor and classmates.
- Highlight key aspects and results of your project and discuss any challenges you encountered

Details:

Zoom Instructions:

We will take a poll in class to decide whether to have presentations in class, online, or hybrid. Regardless, all presentations must be delivered by sharing Zoom screens so that I can record the presentation for evaluation.

Please observe the following for your presentations:

- All team members must present, either in person or online. If online, presenters must have their webcams on during their presentations. The microphones need to be muted, except for the member presenting.
- I recommend having one member display and manage the slides, but each team is free to organize this as you wish.
- All presenters are required to be at a table or desk with proper lighting and audio/visuals or up-front in the classroom. Presenting from a cell phone or with bad lighting and improper setup is not acceptable.
- If you can't secure a convenient and quiet place to present online, then you must present in the classroom.
- The attire is business casual.

- All presentation slides are due no later than 5:15 PM on the day of the presentation (I need a bit of time to download and organize the slides). The project reports are due at midnight the day of the presentation.

Each team will have approx. 10 minutes (including Q&A) to present the project to the class your: business question/problem; model selection; and conclusions. All presentations must follow this format (approximately one slide for each):

- Title slide with the project name and team members' names.
- Business problem and case and business and analytics questions addressed in the study.
- Brief description of the dataset (describe any relevant aspects of descriptive statistics, correlations, visual plot inspections, and pre-processing or transformations, as appropriate).
- Brief explanation of your model selection process and alternatives, along with the respective model specifications.
- Discussion of the final model and the most relevant results. No need to discuss all the results (there is not enough time), just the important ones.
- Final conclusions about the implications of your findings. Are there recommendations for client action that you can make based on the data?
- Brief articulation of the challenges you encountered in your project. Discussing the lessons learned and actionable steps informed by your data insights is a useful way to link your work back to the client, the story, and the strategy. As we stated earlier, effectively communicating data insights to solve a business case is why you as data analysts are in such demand in the professional world.