**Name: Ben Karabinus**

**Question 1: Presentation checklist**

You’ll be giving a presentation next week, so I want you to run through the presentation rubric to ensure that you address all the criteria from the rubric. Please answer the guided rubric questions below to the best of your ability. It’s okay if you make changes to your presentation after you’ve submitted this; in fact, that’s the intention of this exercise. Also, although all the rubric categories will be applicable, not every sub-question about the categories will be applicable to your project; provide answers to the sub-questions as appropriate for your project.

Criteria 1: Data source and definitions explained

**Q1.1: Where did you obtain your data set? Who collected it originally? If you know, what was the motivation for collecting the data? How many variables and observations are in the data set? Which variables are most important to explain to the audience?**

**The data set was obtained from the from a combination of sources. Home values for metropolitan statistical areas in the U.S were obtained from the Zillow Home Value Index (ZHVI) available for download from Zillow.com (**[**https://www.zillow.com/research/data/**](https://www.zillow.com/research/data/)**). Regional price parities, per capita personal income and population for a given metropolitan statistical area in the U.S. were obtained from the U.S. Bureau of Economic Analysis website (**[**https://www.bea.gov/**](https://www.bea.gov/)**). Unemployment rates for metropolitan statistical areas were gathered from the U.S Census Bureau (**[**Census.gov**](https://www.census.gov/)**)**

Criteria 2: Appropriate exploratory data analysis performed

**Q1.2: What kind of exploratory analyses/descriptives are appropriate for your data? What results of your exploratory/descriptive analytic work is most important to convey to the audience to facilitate their understanding of the data in support of the project topic?**

**Appropriate exploratory analysis and descriptive statistics for the data set include measures of central tendency and visualization of distribution of numerical data.**

Criteria 3: Research question presented

**Q1.3: What question does your primary method (i.e., your topic) answer with regard to your data?**

Can the XGBoost Regression algorithm be used to predict Zillow Home Value Estimates given the metropolitan statistical area where a home is located, regional price parity for the metropolitan statistical area, unemployment rate for the metropolitan statistical, per capita personal income for the metropolitan statistical area, population for the metropolitan statistical area and size of the home.

Criteria 4: Method for addressing research question explained

**Q1.4: What is your analysis plan to use your data to answer your research question? What parts of this plan are most important to share with your audience?**

Criteria 5: Primary method explained in principle

**Q1.5: If you had to explain your primary method to a classmate in writing but only had one page to do so, what information would you make sure to put on the page? Put differently, you won’t be able to fully explain any method during your presentation, but what aspects of your primary method must your audience know to understand your project?**

* Gradient descent is the basis for gradient boosted machines, brief explanation of the concept is warranted.
* Compare the more traditional ensemble model building approach to the approach of gradient boosted machines, brief explanation.

Criteria 6: Application of primary method explained

**Q1.6: Having heard your explanation of the primary method**, **would your audience understand your explanation of how you applied the method to your data?**

After hearing my explanation of the primary method I feel the audience would understand my explanation of how I applied the method to my data.

Criteria 7: Data satisfaction of requirements of method demonstrated

**Q1.7: Did you articulate to your audience what properties the data must have in order to use your primary method (e.g., logistic regression requires a binary outcome...) and how your data meet these requirements (e.g., ...and the outcome of interest is binary)?**

Yes, XGBoost for regression requires a continuous outcome and not much else. Data preparation can improve model performance and this will be highlighted in my presentation.

Criteria 8: Primary new method applied and interpreted correctly

**Q1.8: Is your primary method applied in a way that is appropriate for both the data and the research question? Is your interpretation of the results appropriate given your findings and the limitations of your method?**

Yes, the primary method is applied and interpreted correctly. The results will show that even as robust as XGBoost is some data preparation is warranted (i.e., scale and center of highly skewed data with outliers).

Criteria 9: Supporting visualizations provided

**Q1.9: Is there at least one visualization of your data or your analysis results? Do all visualizations support the audience’s understanding of the project? Can each visualization be at least partially interpreted without explanation from the speaker (e.g., Do axes have labels?)**

Two visualizations of the results are provided. Both visualizations have labeled axes and can be interpreted with a very in-depth explanation.

Criteria 10: Presentation style shows sufficient preparation in organization and familiar with topics addressed

**Q1.10: Are you certain you can finish your presentation within the time requirements (5-7 minutes/person)? Have you checked your presentation for spelling/grammar errors? Are your text and pictures readable to the audience? Would your audience believe that you had practiced your presentation at least once or twice before?**

After practicing the presentation, it will be tough. Explaining GBMs in 5-7 to seven minutes is harder than writing the code to create the model.