**Individual Peer Evaluation Form**

Your name: **Jake Meyer**

Write the name of your classmate you are preparing this review for in the designated column. Using a scale of 1-4 (1=strongly disagree; 2=disagree; 3=agree; 4=strongly agree) answer each question. If you aren’t able to answer the question based on what is posted in the discussion board, reach out to your classmate for more information via the discussion board. Total the numbers in each column. **Make sure to answer the questions on the 2nd page.**

|  |  |
| --- | --- |
| Evaluation Criteria | Peer Name: Sucharitha Puppala |
| Has plan in place to complete course project. | **4** |
| Has found datasets/data sources to support project idea. | **4** |
| Has solidified project idea. | **4** |
| Has identified resources for project. | **4** |
| Topic is related to data science and demonstrates topics learned to date through program. | **4** |
| Risks and potential issues have been identified. | **4** |
| TOTALS | **24** |

Feedback on Individual’s project topic:

**The Milestone 2 is well organized and easy to follow. The abstract details what will be discussed (i.e. an introduction to Predictive Analytics, the dataset, along with the models that will be used for the project). The Predictive Analytics section provides a great overview of the topic and list potential benefits. One initial piece of feedback to consider is the audience the paper will be geared towards. If this paper were submitted to a manager or stakeholders, then it makes sense to include the benefits of Predictive Analytics along with an overview of the subject. This helps support the positive cases for moving forward with this project. However, if this paper were submitted to colleagues, then it can be assumed that the reader will have a good understanding of Predictive Analytics prior to reading the paper. This is just something to consider and I don’t believe this section needs to be removed since this Milestone focuses on the business proposal (considered more along the lines of manager audience). Next, the introduction provides great background for the topic focused on the India rental housing market. There is clearly value that would come from this project to predict house rent availability around various locations of India. The dataset is distinctly overviewed along with access provided with a link to the Kaggle source. The size of the dataset seems adequate with over 4700 records available for the various living options across twelve different features. The models that will be considered for this project are Linear and Logistic Regression. These two predictive models will be utilized to address specific problems. Another piece of feedback is to specifically call out which problem each model will address. This will help the reader key in on the predictive task(s) for Linear and Logistic Regression. The plan for evaluation seems sound with the Train/Test method along with metrics such as R-Squared and confusion matrices for prediction accuracy and relevance. For the Scope section, the items you hope to learn are identified. In addition to personal learnings for this project, it may be beneficial to include objectives for learnings specifically for the project. The last three sections consist of risks, a contingency plan, and additional elements. Uncertainty within a model is a risk to consider and important to document if the model is deployed. The contingency plan outlines the need to find alternative data if this dataset does not work out as intended. Very thorough outline Sucharitha, great job on Milestone 2!**

1. How clear is the classmate’s project topic? What questions does their topic make you consider?

**The project topic is outlined very clearly. The project will focus on prediction of rent around various locations of India. This topic raises the following questions:**

* **What are the main contributors for predicting rent cost?**
* **What are the main contributors for predicting rent availability?**
* **Which locations are preferred for renting in India?**
* **Which locations are predicted to have increased rent?**

1. What risks or issues should your classmate consider while working on their project?

**The dataset appears to have a mix of categorical and numeric data. For some of the categorical data, caution should be taken in the data understanding phase to ensure the various categories make sense. For example, the Floor, Area Locality, and City may have slight variations in text for categories that are the same. There may be a space before or after some of the text, capitalization differences, or even differences in spelling. The categorical data should also be checked for balance for the various categories. The numerical data, especially the target variable, needs to be understood for how the data is distributed. Check for outlier concerns during the data understanding phase as well. It appears that you have a good contingency plan to find alternative data if this dataset does not turn out as expected.**

1. Additional suggestions/comments that might be beneficial to your peer?

**You have a great plan outlined for the Course Project. For the references, I would recommend following the APA in-line citation format for the resources referenced within the paper. This will help the reader identify which statements are tied to which resource. My only additional suggestion would be to create a list of questions you hope to answer or learn more about specifically related to the India rent topic. These questions can be addressed during the EDA (Exploratory Data Analysis) phase, the Predictive Model phase, or at any point during the analysis. This may help stay on path for questions throughout the analysis. Overall, great job and it appears you are on track!**

Adapted from a peer evaluation form developed at Johns Hopkins University (October, 2006)