Individual Peer Evaluation Form

Your name: Blandon Lee

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:
Has plan in place to complete course project.	4
Has found datasets/data sources to support project idea.	4
Has solidified project idea.	3
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.	3
TOTALS	22

Feedback on Individual's project topic:

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

Blandon is trying to predict who can police shootings be avoided in the future based on the data he downloaded from Kaggle. He has a plan to create a model to predict shootings, find out the locations for higher death rates per demographics, and examine how age, location, and race impact shootings. Therefore, he clearly understood what he needs to accomplish to get his result. I believe he has correctly chosen the K-modes clustering model to define to meet the object for his assignment because of the categorical data he has chosen. I think Linear Regression and Decision Trees are the appropriate models for his analysis. I would consider asking, How many people are killed by the police annually?

The Washington Post says, half of the people shot and killed by police are white. Black people count for less than 13 percent of the U.S population but are killed by police at more than twice the rate of white Americans. Is it possible to find the motive for the killing of black people by policemen in this analysis?

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

I believe Blandon understands what model he would like to use, but the main risk with this proposal is what is that plan if the result is inconclusive?

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

Do you have to decide which evaluation metric is the best for your project?

The K-modes algorithm accepts np. NaN values as missing values. However, I would suggest that to consider filling in the missing data themselves in a way that makes sense for the problem at the hand.

You don't need to do correlation analysis as the target variable because of categorical variables.

Another useful visualization is a barplot of the distribution of the target variables.

Random Forest Classification model allows you to access important features.