**Audience Questions:**

1. **How realistic is it to apply a simulated dataset to real-world placement strategies?**

While simulated datasets do not show actual real world data, they can still be a practical starting point for model development. Especially when real data is scarce or confidential. However, while these datasets can replicate real patterns, simulated data may lack some of the noise or diversity of actual student populations. These models should be used as prototypes or a baseline and then retrained with real-world data from each university.

1. **What is the main advantage of using machine learning over traditional statistical methods for this problem?**

Machine learning models can automatically capture complex, non-linear relationships and interactions between variables without needed specification. This is especially true for tree-based algorithms like the Random Forest model we used here. This flexibility often results in higher accuracy scores compared to other models.

1. **Why were certain features like gender or socioeconomic status not included in the dataset?**

These features may have been excluded to avoid potential ethical issues. Including these features might raise legal concerns if used in high-stakes decisions. The goal is to focus on merit based and performance driven indicators.

1. **Was there a feature that you expected to be influential that wasn’t?**

Yes, there were a few results that were particularly surprising. Features like the extracurricular involvement may be expected to correlate with placement, but may not show strong positive correlation. This could be due to a number of things including limited variance in the data or employers prioritizing technical skills.

1. **Which feature turned out to be the most predictive of placement? Why?**

Performance related features turned out to be the most predictive. These include things like degree percentage or professional exam scores. These variables directly reflect a student’s aptitude and may be more closely aligned with their employers expectations.

1. **How can you address potential class imbalance in placement outcomes?**

There are several ways these potential imbalances could be addressed. The first would be resampling. Another would be to use models that account for class imbalance. The third could be to apply a metric adjustment.

1. **What steps were taken to validate the model’s accuracy?**

The model was validated using a train-test split, a confusion matrix analysis, and classification metrics. These measures ensured the model performed consistently across unseen data.

1. **How would you ensure transparency and fairness if this model were deployed institutionally?**

Regular audits on this data and model would give the opportunity to check for biases in the predictions. As stated previously, excluding protected features could help keep students records anonymous, but also limit the likelihood that bias could take place.

1. **What are some specific actions colleges could take based on the model’s results?**

Colleges could use this model to identify students who are at risk of not being placed in a role out of college and provide them with resources and counseling to increase their chances. They could also adjust curriculum emphasis in areas that are low-performing for many students.

1. **What are some ways to collect additional real-world data to improve predictions?**

Colleges could collect their own data from the students at their school. Tracking alumni career paths could help apply lessons from others experiences to students actively enrolled in the school.