1. **What does the CRISP-DM methodology for data mining entail? What does the data analyst do in each step?**

First, let's talk about CRISP-DM, the Cross Industry Standard Process for Data Mining. It's a comprehensive methodology that involves six phases: understanding the business problem, understanding the data, preparing the data, modeling, evaluating, and finally deploying the solution. In each phase, the data analyst works closely with the data and stakeholders to ensure that the final model not only solves the business problem but also performs optimally. It's an iterative process where the data analyst is constantly refining both the data and the model to ensure it meets business objectives.

2. **Explain what needs to be done before you can forecast time series. In other words, what does time series analysis involve?**

Next, before forecasting time series data, it's crucial to understand that time series analysis involves identifying patterns in the data, such as trends, seasonality, and cyclical behavior. To accurately forecast future values, the data needs to be stationary, which means its statistical properties don't change over time. This often requires transformation techniques like differencing or detrending. Once the data is prepared, the analyst can then apply forecasting models like ARIMA or exponential smoothing.

3. **Discuss the main principles of data ethics, as well as what companies or organizations could do to uphold these principles.**

Data ethics is another vital area in our work. The main principles include fairness, transparency, and accountability. Companies need to ensure that data is collected and used responsibly, protecting individuals' privacy and avoiding harm. This involves being transparent about data use, ensuring that algorithms are free from bias, and holding themselves accountable for any misuse of data. Upholding these principles is not just about compliance—it's about maintaining trust with customers and stakeholders.

4. **Imagine you're an analyst at a financial services company and you have some concerns about bias in the way a predictive model was measured. How would you raise your concerns with your manager? Make some suggestions for how to avoid measurement bias in the future.**

Now, let's consider the topic of bias in predictive models. If I were concerned about bias in the way a model's performance was measured, I would raise these concerns with my manager by explaining how certain metrics or data points might skew the results. I would suggest conducting tests using more diverse datasets and implementing fairness checks throughout the model-

building process. Avoiding measurement bias can be achieved by using more balanced training data and regularly evaluating model outputs for any signs of unfair treatment.

Explain the difference between regression and classification models in predictive analysis. When would you use one of these models over the other?

Finally, let's clarify the difference between **regression** and **classification models** in predictive analysis. Regression models are used when we're predicting a continuous numerical value, like stock prices or sales figures, whereas classification models are used to categorize data into groups, such as whether a transaction is fraudulent or not. The choice of model depends on the problem—use regression when the output is a number, and classification when it involves a label or category.