Q1)

Our topic will be focused on predicting whether mushrooms are poisonous or safe to consume.

Q2)

1. This problem will be a classification problem.
2. Eating a poisonous mushroom could potentially lead to a life-threatening situation. As a result, we would want our model to minimize the false negative rate (FNR), i.e., minimize the probability of predicting a safe-to-eat mushroom and it turning out to be poisonous. This will include optimizing for precision and recall as well.
3. N/A
4. We are aiming for as close to 100% accuracy as possible, as the stakes of our prediction are extremely high.
5. Mushrooms are highly complex organisms, having a wide variety of patterns and physical attributes. This leads us to believe that a non-linear model would be most helpful in predicting whether or not they are poisonous.

Q3)

The data is from Kaggle and includes the classification variables class, cap-shape, cap-surface, cap-color, bruises, odor, gill-attachment, gill-spacing, gill-size, and gill-color. Each observation lists whether a mushroom is poisonous or not alongside the listed variables. From the data, we can compose a classification model to determine if any mushroom is poisonous or not based on its physical properties.

https://www.kaggle.com/datasets/uciml/mushroom-classification

Q4)

Our stakeholders will mostly consist of naturalists looking to provide wilderness insights and warnings to others. Other stakeholders could include hikers, mycologists, or chefs looking to use fungi in their cuisine.

Q5)

We expect to develop with great accuracy a model to predict and classify which mushrooms are poisonous. As an additional task, we will look to keep simplicity of the model in mind. This will increase the generalizability to stakeholders as they could better classify independently with a simpler set of guidelines.

Q6)

Jack: data cleaning and preparation, initial data analysis

Wilson: data visualization, base model

Sahil: model selection

Glen: model refining, report structure