Modeling Instructions

Follow the steps below to guide you through the modeling process:

1. **Prepare Features and Target Variables:**
    - Create a feature matrix `X` and target variable `y`. Ensure that `X` does not include columns such as respondent_id and the target variable (price_range).

2. **Data Splitting:**
    - Split the dataset into training and test sets, using 75% of the data for training and 25% for testing. In train_test_split function call, please use random_state value of 42, this way your notebook and our notebook have same split and it helps with results verification

3. **Feature Encoding:**
    - Apply appropriate encoding techniques to the features:
        - Use Label Encoding for the following columns: age_group, income_levels, health_concerns, consume_frequency(weekly), and preferable_consumption_size.
        - Apply One-Hot Encoding to all remaining categorical columns.
        - Ensure the target variable (price_range) is also label encoded.

4. **Model Selection:**
    - Test the following machine learning models on the prepared data:
        - Gaussian Naive Bayes
        - Logistic Regression
        - Support Vector Machine (SVM)
        - Random Forest
        - XGBoost
        - Light Gradient Boosting Machine (Light GBM)

5. **Performance Evaluation:**
    - For each model, calculate and print the accuracy score and the classification report.

6. **Model Comparison:**
    - Track the performance of each model and select the best-performing model for the next steps (which will be communicated in future tasks).

If you encounter any difficulties during the modeling phase, don't hesitate to reach out to your senior team members for guidance.
