Submit a proposal for a recipe classifier for different cuisine types.

- Data collection
    - To correctly classify different recipes, we would need to gather and organize a large number of food types and instructions on how they are made. This can be done either through web scraping certain websites, public datasets, and even homemade recipes.
    - Filter posts or comments containing recipes based on specific keywords or tags related to different cuisines.

- Data exploration & preprocessing
    - Cleaning the data is just as important as gathering the data. It is necessary to make sure that all the information is formatted correctly and that there are no missing or unusable variables.
    - Normalize text data by tokenizing
    - Understanding class/label distribution: balanced or imbalanced

- Annotation
    - Set up Prodigy to annotate the dataset with cuisine labels.
    - Develop annotation guidelines and discuss edge cases such as “fusions” and “creoles” and where they fit in our labels, perhaps adding more labels.
    - Each of us annotators would manually label a subset via Prodigy interface
    - Prioritize and flag annotations of recipes of uncertain or ambiguous examples for efficiency
    - Continuously update and refine annotation guidelines based on annotator feedback to ensure consistency in labeling over time and between annotators.

- Training and evaluation
    - We would need to categorize the types of cuisine based on different details such as ingredients and the cooking methods used to create the dish. Important to note that both ingredients and the method used are not exclusive and can apply to many different cuisine types.
    - Regional and cultural differences need to be discussed and written in the annotation guidelines

- Deployment
    - Deploy the trained model as an API or similar service to classify new recipes into cuisine types.
    - Develop a user-friendly interface to allow users to input recipes and receive predicted cuisine labels.
    - Monitor and manage model: Ensure scalability and reliability of the model itself and the system to accommodate new data and varying levels of user traffic.