Step 1: Preprocessing. Chat is in a text file, need to extract this text and convert it into pandas dataframes