import re  
import pandas as pd  
def preprocess (data):  
 pattern = '\d{1,2}/\d{1,2}/\d{2,4},\s\d{1,2}:\d{2}\s-\s'  
 messages = re.split(pattern, data)[1:]  
 dates = re.findall(pattern, data)  
 df = pd.DataFrame({"user\_message": messages, "message\_date": dates})  
 df["message\_date"] = pd.to\_datetime(df['message\_date'], format="%d/%m/%y, %H:%M - ")  
 df.rename(columns={'message\_date': 'date'}, inplace=True)  
  
 users = []  
 messages = []  
  
 for message in df['user\_message']:  
 entry = re.split('([\w\W]+?):\s', message)  
 if entry[1:]:  
 users.append(entry[1])  
 messages.append(" ".join(entry[2:]))  
 else:  
 users.append('group notification')  
 messages.append(entry[0])  
 df['user'] = users  
 df['message'] = messages  
 df.drop(columns=['user\_message'], inplace=True)  
  
 df['year'] = df['date'].dt.year  
 df['month\_num'] = df['date'].dt.month  
 df['month'] = df['date'].dt.month\_name()  
 df['day'] = df['date'].dt.day  
 df['only\_date'] = df['date'].dt.date  
 df['day\_name'] = df['date'].dt.day\_name()  
 df['hour'] = df['date'].dt.hour  
 df['minute'] = df['date'].dt.minute  
  
 period = []  
 for hour in df[['day\_name', 'hour']]['hour']:  
 if hour == 23:  
 period.append(str(hour) + "-" + str('00'))  
 elif hour == 0:  
 period.append(str('00') + "-" + str(hour + 1))  
 else:  
 period.append(str(hour) + "-" + str(hour + 1))  
  
 df['period'] = period  
  
  
 return df