**CODE for Upper CASE**

**import pandas as pd**

**import numpy as np**

**import matplotlib.pyplot as plt**

**import seaborn as sns**

**# Function to load cleaned subject data**

**def load\_cleaned\_subject\_data(file):**

**return pd.read\_csv(file)**

**# List of cleaned subject files**

**cleaned\_subject\_files = ['subject-6.csv']**

**#'subject-3.csv','subject-5.csv','subject-6.csv','subject-7.csv','subject-8.csv']**

**# Process each subject's data**

**for file in cleaned\_subject\_files:**

**subject\_id = file.split('.')[0] # Extract subject ID from the file name**

**df\_subject = load\_cleaned\_subject\_data(file)**

**# df\_subject = df\_subject.dropna() # Remove rows with NaN values**

**# Keep only the last occurrence of each FPOGID in the entire dataframe**

**#df\_subject\_duplicate = df\_subject.drop\_duplicates(subset='FPOGID', keep='last')**

**print(df\_subject.columns)**

**df\_upparcasestart=df\_subject['USER']=='SENTENCE\_UPPERCASE\_ONSET'**

**df\_upparcaseend=df\_subject['USER']=='UPPERCASE\_DIALOGUE\_END'**

**# Initializing an empty list to store the extracted segments**

**extracted\_segments = []**

**# Iterating over the DataFrame to extract segments between start and end markers**

**start\_index = None**

**for index, row in df\_subject.iterrows():**

**if df\_upparcasestart[index]:**

**start\_index = index**

**elif df\_upparcaseend[index] and start\_index is not None:**

**end\_index = index**

**extracted\_segments.append(df\_subject.loc[start\_index:end\_index])**

**start\_index = None**

**# Concatenate all extracted segments into a single DataFrame**

**df\_extracted = pd.concat(extracted\_segments, ignore\_index=True)**

**df\_subject\_duplicate = df\_extracted.drop\_duplicates(subset='FPOGID', keep='last')**

**sns.displot(df\_subject\_duplicate, x="FPOGD")**

**plt.title(f'{subject\_id}- uppercase')**

**plt.ylabel('Count')**

**plt.xlabel('Fixation duration in s')**

**plt.ylim(0, 25)**

**plt.show()**