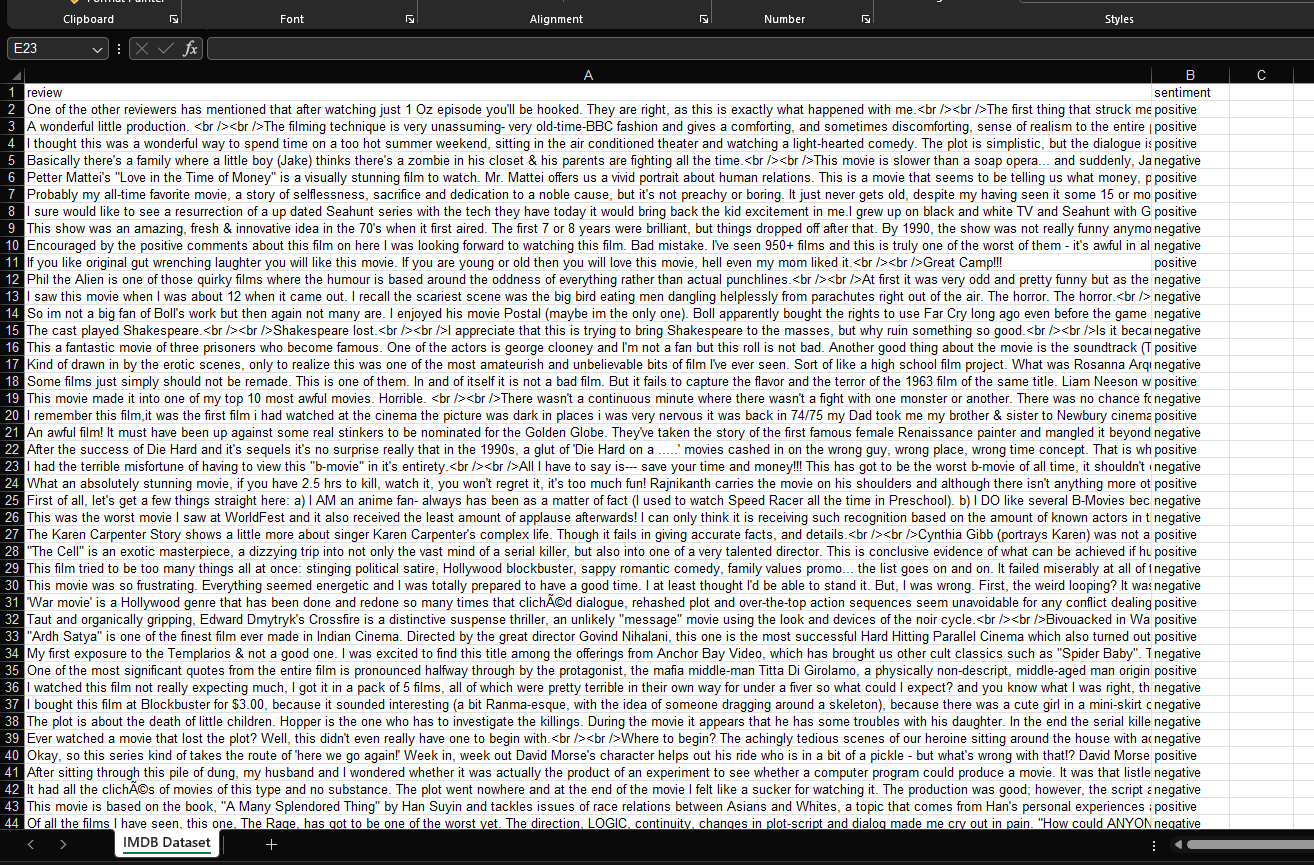
**Group 8 – Project Dataset Proposal**

**Ramon Castro, Cintia Hernandez, Jhose Valderrama & Madeline Mesa**

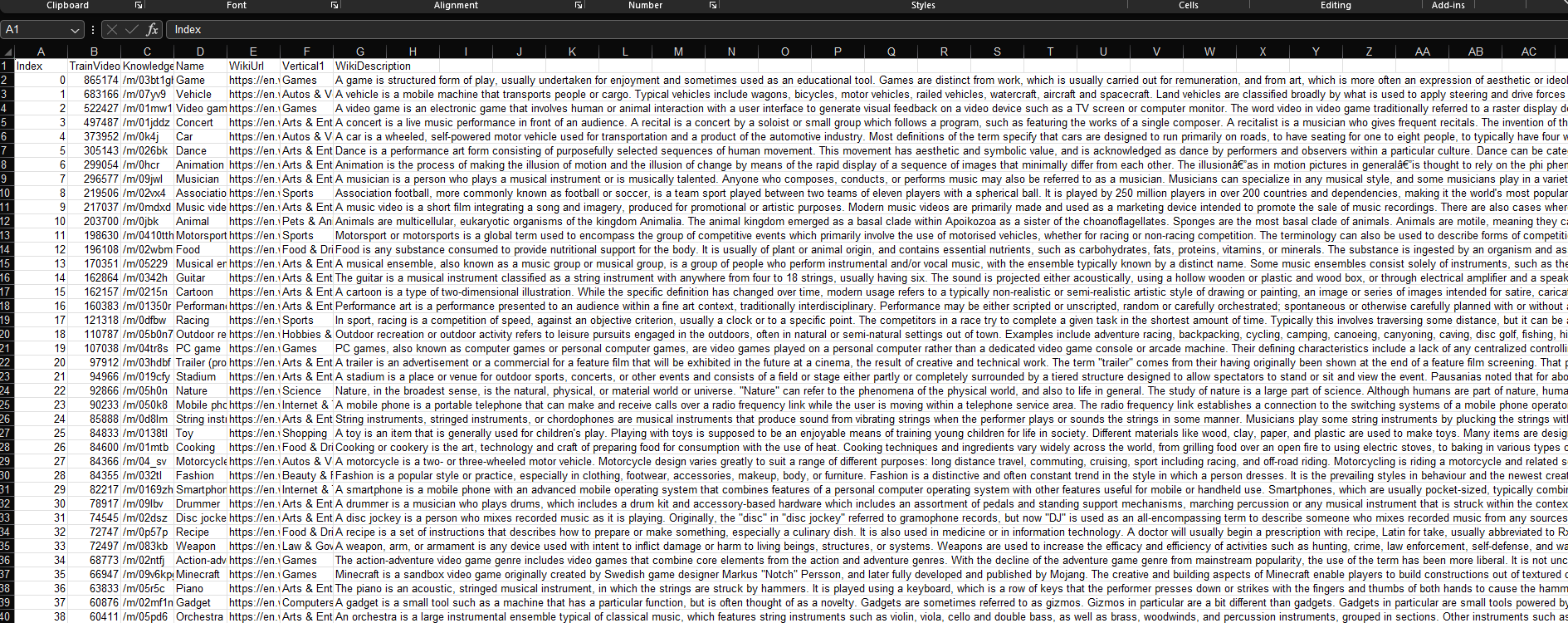
**Project Proposal - Automating Video Editing**

1. Audience Engagement Data
   * Acquire audience ratings, comments, and engagement data for Movies from different sources.
   * IMBD Reviews Sentiment Analysis
     + This dataset contains the features 'review' and 'sentiment'.  With this information we can complete a sentiment analysis and train a model on customer engagement.  Training the Ai on user preference will allow us to tailor the editing style and process.
     + <https://www.kaggle.com/datasets/lakshmi25npathi/imdb-dataset-of-50k-movie-reviews>
   * YouTube 8M
     + This dataset contains millions of YouTube video IDs, and it is already labeled.  It can be useful for analyzing and categorizing video data.  For example, we can categorize videos into genres.  This will allow us to potentially train the Ai to use a specific tone, style or color when editing an action movie versus a romance movie.
     + <https://research.google.com/youtube8m/>
2. Facial and Object Recognition Data
   * Acquire datasets with images and videos for object detection and facial recognition.
   * Hollywood2 Action Dataset
     + This dataset provides human action recognition in different environments.  This can be useful to train a model to recognize a scene based on the character’s action.  The Ai can determine the editing approach based on this.
     + https://www.di.ens.fr/~laptev/actions/
   * Labeled Faces in the Wild - LFW
     + This dataset can help us with further facial recognition.  It helps in the automation of identifying and tagging characters of a movie within a scene.
     + https://vis-www.cs.umass.edu/lfw/

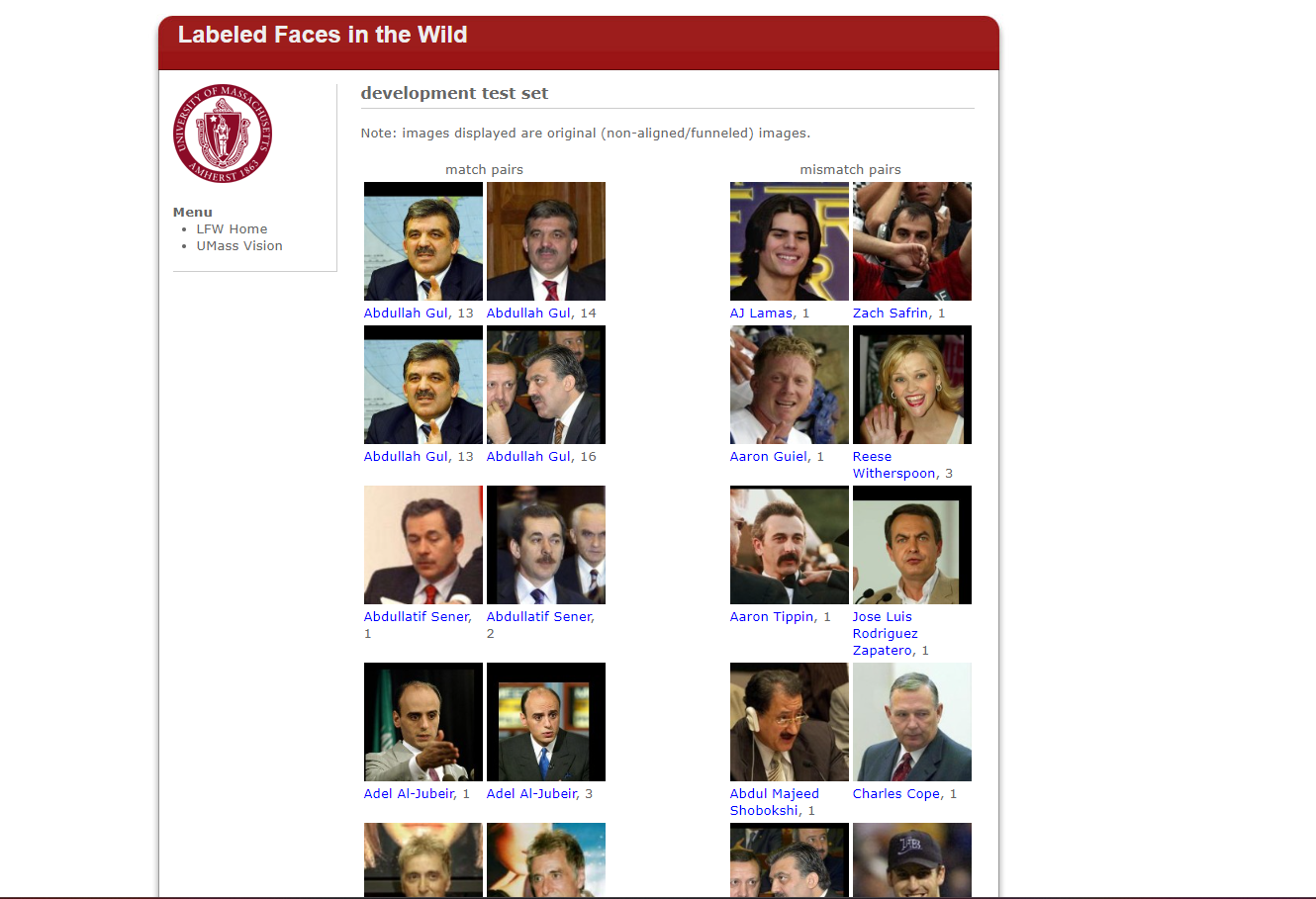
IMBD DATASET



YOUTUBE 8M DATASET



LABEDLED FACES IN THE WILD DATASET



HOLLYWOOD2 ACTION DATASET

