Exploratory Data Analysis

REPORT

**1. Project Overview**

**Objective**

The main goal of this analysis is to evaluate the performance of candidates based on their introduction videos by analyzing various metrics, including emotion scores, transcript scores, and other visual features. This evaluation aims to generate actionable insights for recruitment by integrating emotion analysis with speech and visual cues.

**Datasets Used**

The analysis is based on two primary datasets:

1. **Emotion Data (CSV file):** Contains detailed scores for various emotions and other related metrics. Emotional metrics like angry, happy, sad, neutral, etc., over time for each candidate.

* **Gaze Data**: Provides information about the candidate’s eye movement and blink rates, which can be used to assess engagement.

1. **Transcript Data (CSV file):** Contains transcripts of the candidates' speeches, including additional scoring data related to the speech content.
2. **Transcript Data (text file): Contains the text of the transcript in a** paragraph of the candidate throughout the video.

**Features in the Datasets**

* **Emotion Data:**
  + movie\_id: Unique identifier for each video.
  + image\_seq: Sequence number of images/frames.
  + emotion\_scores: Various scores indicating the intensity of emotions, such as angry, disgust, fear, happy, sad, surprise, and neutral.
  + dominant\_emotion: The most prominent emotion detected in each frame.
  + gaze: Indicates whether the candidate is looking at the camera (1) or not (0).
  + blink: Indicates whether the candidate blinked (1) or not (0).
  + eye\_offset: Measurement of deviation from the camera.
  + elapsed\_time: Timestamp in seconds.
* **Transcript Data:**
  + movie\_id: Matches with the emotion data.
  + start and end: Start and end times of speech segments.
  + text: Transcript text.
  + positive, negative, neutral: Sentiment scores of the speech.
  + confident, hesitant, concise, enthusiastic: Scoring metrics related to speech delivery.
  + speech\_speed: Speed of speech delivery.

**2. Data Preprocessing**

For each candidate, the following preprocessing steps were performed:

1. **Alignment of Transcript and Metadata**:
   * The transcript data (with start and end times) was aligned with the metadata, allowing us to merge time-based data such as emotions and gaze.
2. **Data Cleaning**:
   * Missing or erroneous data points were handled, and numerical columns were normalized or scaled as needed.

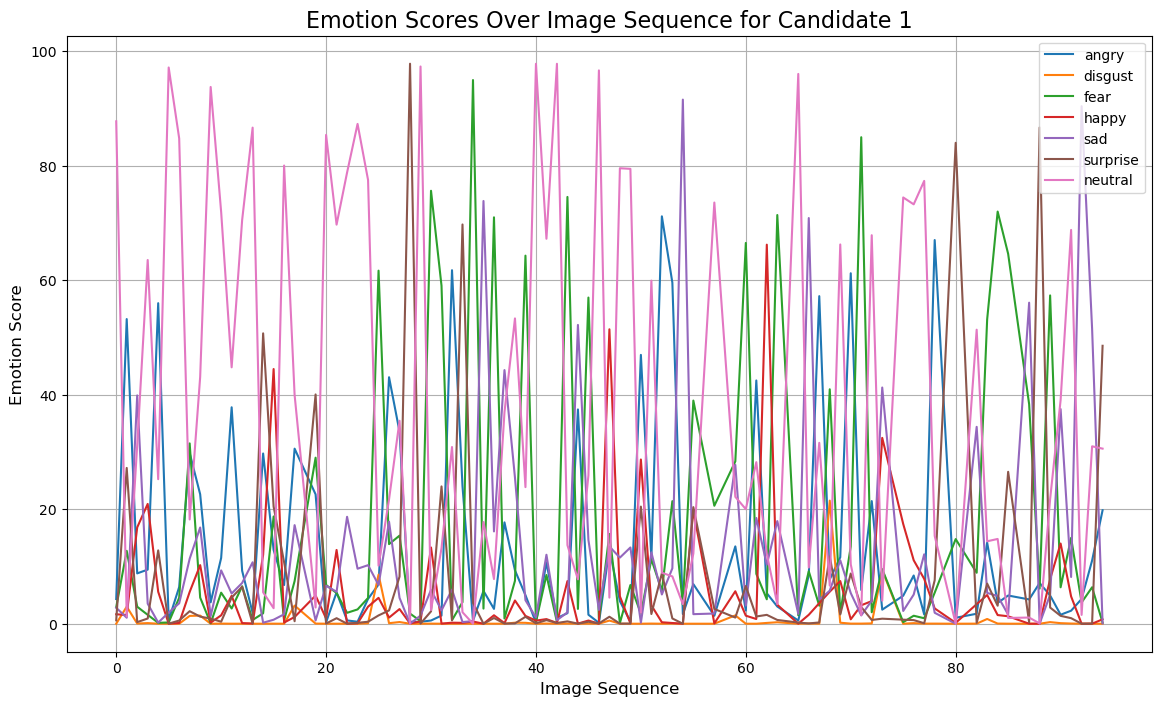
**3. Visualizations**

**3.1. Emotion Data Analysis**

For each candidate, the following emotional metrics were plotted over time:

* **Angry**
* **Happy**
* **Sad**
* **Fear**
* **Neutral**
* **Surprise**

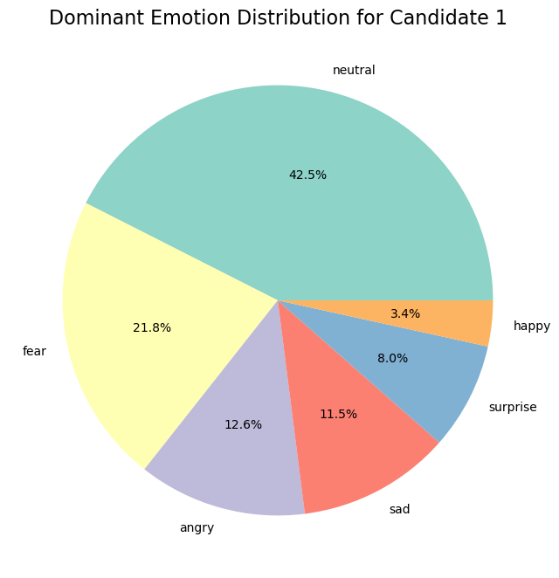
**Example Visualization: Emotions Over Time for Candidate 1**

* 

This plot shows how each emotion fluctuates during the candidate’s speech, offering insights into their emotional state at different times.

**3.2. Dominant Emotion Distribution**

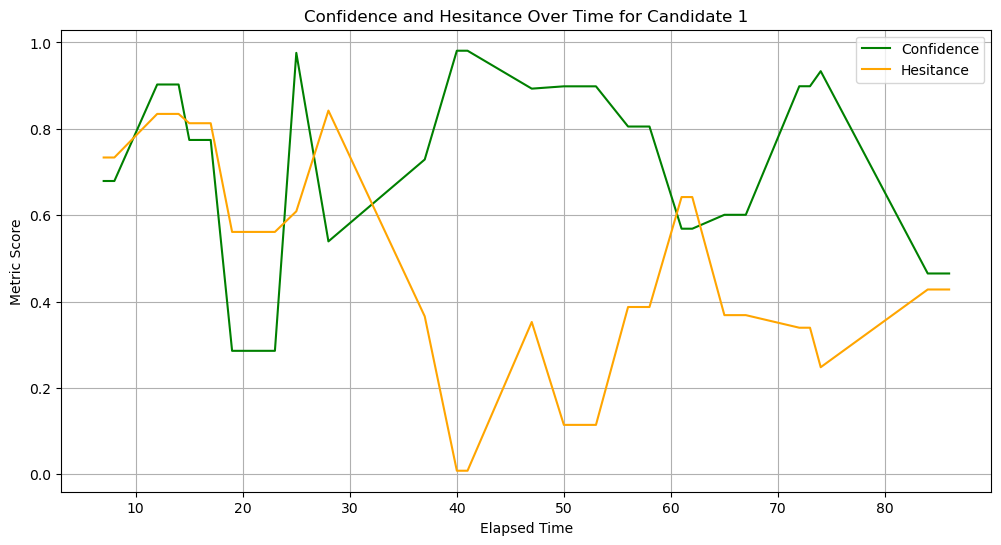
Each candidate's dominant emotion was extracted from the data, and a pie chart was generated to visualize the distribution.

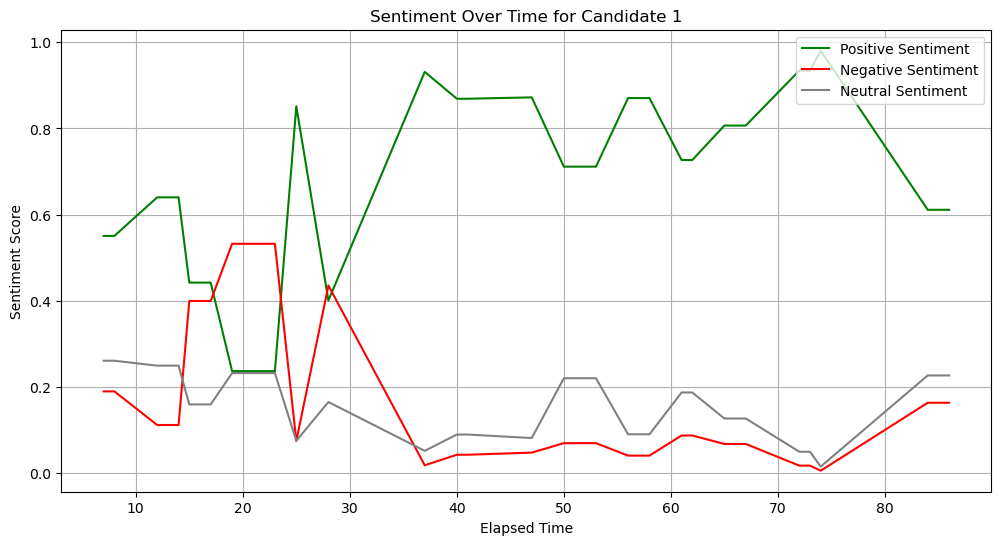
**Example Pie Chart: Dominant Emotion Distribution for Candidate 1**

**3.3. Communication Metrics**

Over time, we visualized key communication metrics such as confidence and hesitance. This helped us observe how fluid and confident each candidate's speech was.

**Example Visualization: Confidence and Hesitance and sentiments Over Time**

* 



**4. Candidate Scoring**

After the visualization, we calculated various scores for each candidate, focusing on the following metrics:

* **Confidence**: How confident the candidate's speech was.
* **Hesitance**: The level of hesitation observed in speech.
* **Positive Sentiment**: Proportion of positive speech.
* **Emotional Stability**: Based on emotion metrics (neutral and happy states).
* **Engagement**: Measured using gaze data (average gaze and blink rates).

Each candidate's scores were normalized and aggregated into a final score.

**4.1. Ranking Candidates by Score**

The final scores were calculated using a weighted sum of the key metrics:

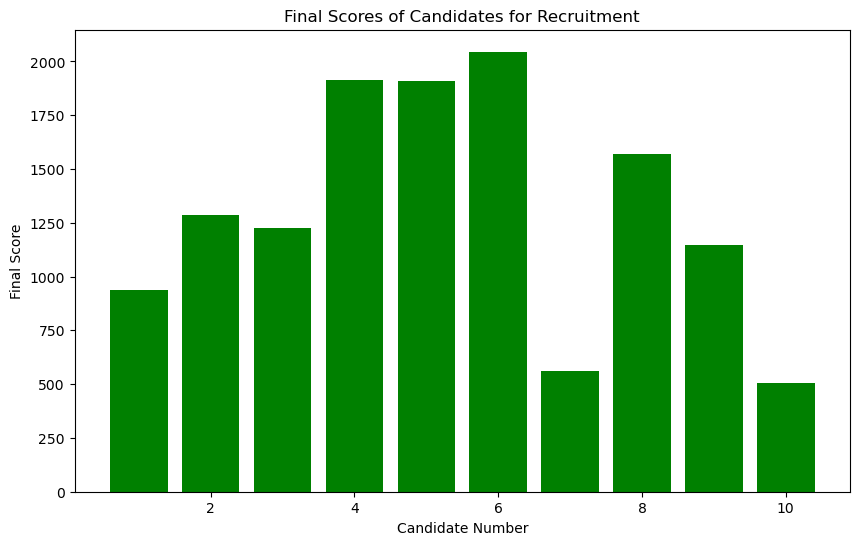
final\_score = (confidence \* 0.3 + hesitance \* 0.2 + positive\_sentiment \* 0.2 + emotional\_stability \* 0.2 + engagement \* 0.1)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Candidate** | **Confidence** | **Hesitance** | **Emotional stability** | **Positive Sentiment** | **Engagement** | **Final Score** |
| 1 | 73.382801 | 51.482753 | 4422.896609 | 70.919907 | 62.500000 | 937.324694 |
| 2 | 68.487934 | 56.384181 | 6163.988848 | 72.200636 | 60.919540 | 1285.153067 |
| 3 | 57.356613 | 39.599574 | 5926.914628 | 56.725685 | 45.000000 | 1226.354961 |
| 4 | 62.174019 | 42.954833 | 9320.159082 | 65.574828 | 78.787879 | 1912.268742 |
| 5 | 59.009398 | 53.851214 | 9287.326428 | 63.057273 | 100.000000 | 1908.549803 |
| 6 | 67.975475 | 50.974773 | 9951.920071 | 71.118230 | 100.000000 | 2045.195257 |
| 7 | 70.371388 | 54.293029 | 2536.821136 | 71.735390 | 78.160920 | 561.497420 |
| 8 | 55.501081 | 49.237767 | 7607.717655 | 60.540213 | 94.623656 | 1569.611817 |
| 9 | 59.184206 | 46.126830 | 5486.586302 | 61.735264 | 96.511628 | 1146.296104 |
| 10 | 61.985230 | 47.936334 | 2296.350252 | 58.926690 | 73.333333 | 506.571558 |

**4.2. Visualizing Final Scores**

A bar plot was generated to compare the candidates' final scores visually:

**Final Scores Bar Plot:**

* 
* Ranking of candidates based on thresholds is:

6 > 4 > 5 > 8 > 2 > 3 > 9 > 1 > 7 > 10

**5. Recruitment Recommendations**

Based on the candidate scores, the following decisions were made:

* confidence\_threshold = 70
* hesitance\_threshold = 50
* positive\_sentiment\_threshold = 60
* emotional\_stability\_threshold = 65
* Candidates exceeding all thresholds were recommended for recruitment.
* Candidates with lower scores in critical areas were flagged for potential improvements.

**6. Communication Analysis**

In addition to the quantitative metrics, qualitative feedback was provided based on the candidate's communication patterns:

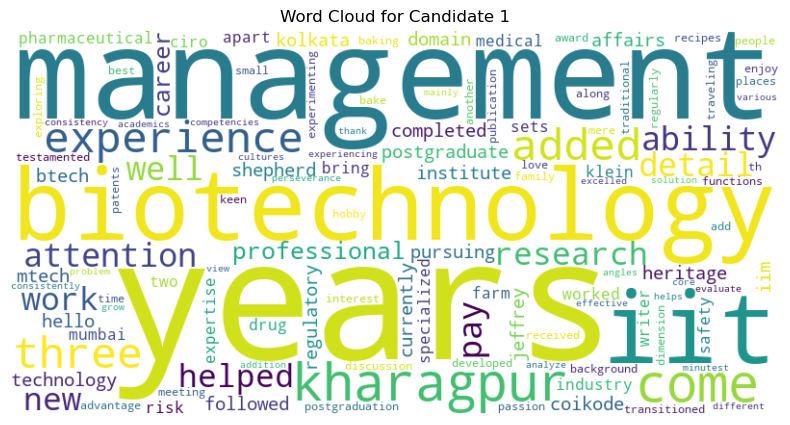
**Example Analysis for Candidate 1:**

* **Confidence**: High confidence in communication.
* **Hesitance**: Smooth and fluent communication.
* **Sentiment**: Positive and engaging tone.
* **Improvement**: Minor improvements in emotional stability could enhance overall performance.

**7. Word Cloud Analysis**

Finally, word clouds were generated for the transcript data to identify key themes and topics for each candidate.

**Example Word Cloud for Candidate 1:**



The word cloud highlights the most frequently used words in each candidate's speech, offering insight into their primary topics and focus areas.

Using the aggregated metrics like confidence, hesitance, positive sentiment, emotional stability, and engagement, we can make data-driven decisions about whether or not to recruit the candidate. These decisions will be based on how well the candidate performs in key areas relevant to the job.

**8. Findings:**

**High Confidence + Low Hesitance:**

These are the leading indicators of effective communication and the ability to present oneself. A high confidence score with low hesitance indicates a strong candidate.

**Positive Sentiment and Emotional Stability:**

Candidates with a generally upbeat tone and stable emotional control (neutral or positive emotions) are preferred, especially for roles requiring consistent performance under pressure.

**Engagement:**

A steady gaze and fewer blinks reflect attentiveness and a well-prepared candidate. This can help gauge a candidate’s engagement level during the interview.

**Key Metrics for Communication Skills:**

Confidence: Measures the candidate’s ability to speak with certainty.

Hesitance: Reflects how smoothly the candidate speaks (lower is better).

Speech Speed: Helps gauge whether the candidate is speaking at a natural pace, avoiding excessive speed or slowness.

Clarity (Conciseness): A high score in conciseness means the candidate can communicate without unnecessary information, an essential trait in many job roles.

Positive Sentiment: A positive tone can indicate enthusiasm and engagement during communication.

**Areas of Expertise Based on Data:**

Highly Confident but Hesitant:

Some candidates might score highly in confidence but still have moments of hesitation. This could indicate that they know the material but may need more practice in public speaking or in front of cameras.

Good Sentiment Control:

A candidate with balanced positive sentiment and neutral emotions may be more suited for roles requiring emotional intelligence and customer interactions.

Speech Speed:

Candidates with optimal speech speed (neither too fast nor too slow) can indicate that they think clearly and articulate well.

**Additional Insights for Recruitment Decisions:**

**Gaze and Blink Patterns:**

Candidates who maintain firm eye contact and have fewer blinks are likely more engaged and less nervous, which are positive traits for job roles that require interpersonal skills or public presentations.

**Emotional Fluctuations:**

Candidates who exhibit frequent emotional changes (e.g., anger or sadness) might struggle under stress or pressure. On the other hand, stable emotions are a sign of emotional control, which is desirable in leadership or customer-facing roles.

Creative Insights:

Candidates with High Positive Sentiment and Neutral Emotions:

These candidates will likely handle stressful situations better while maintaining a calm and professional demeanor. They might be suited for roles requiring high emotional intelligence.

Candidates with High Enthusiasm:

High enthusiasm indicates a passion for the role, which can be valuable in jobs requiring creativity or leadership.

**Summary :**

This report comprehensively analyzes 10 candidates based on emotion, gaze, and transcript data. The findings were visualized through various plots, and a scoring system was developed to rank and recommend candidates for recruitment.

The below link is the access to the code in the Jupyter Notebook :

<https://colab.research.google.com/drive/17Ne3bFSC3KITUOsgO88aXKIVqBuY9Umn>

TEJASRI RELANGI

MS22BTECH11020

IIT HYDERABAD