Exploratory Data Analysis Report

This report presents an exploratory data analysis (EDA) of 10 candidates based on their introduction videos. The primary goal of this analysis is to assess each candidate’s suitability for recruitment, evaluate their communication skills, and identify areas of expertise that can inform hiring decisions.

The analysis is centred around three main data components:

1. **Emotion Scores**: These scores capture the candidates' emotional expressions throughout their videos, providing insights into their engagement, confidence, and overall presence during their presentations. The analysis of emotions such as happiness, neutrality, surprise, and other sentiments helps in understanding how well candidates connect with their audience.
2. **Transcript Scores**: These scores are derived from the content analysis of the candidates' transcripts. The metrics include positivity, confidence, conciseness, and other aspects that reflect the quality and impact of their speech. This evaluation helps determine how effectively candidates communicate their thoughts and present their qualifications.
3. **Text Scores**: These scores are based on the relevance of recruitment-related keywords and the confidence shown in delivering these terms. Keywords like "experience," "leadership," "skills," and "teamwork" are particularly emphasized, as their presence indicates the candidates’ ability to convey relevant professional experience and qualifications.

**Data Summary**

**Dataset Overview:**

* **Number of Candidates**: 10

**Dataset Components:**

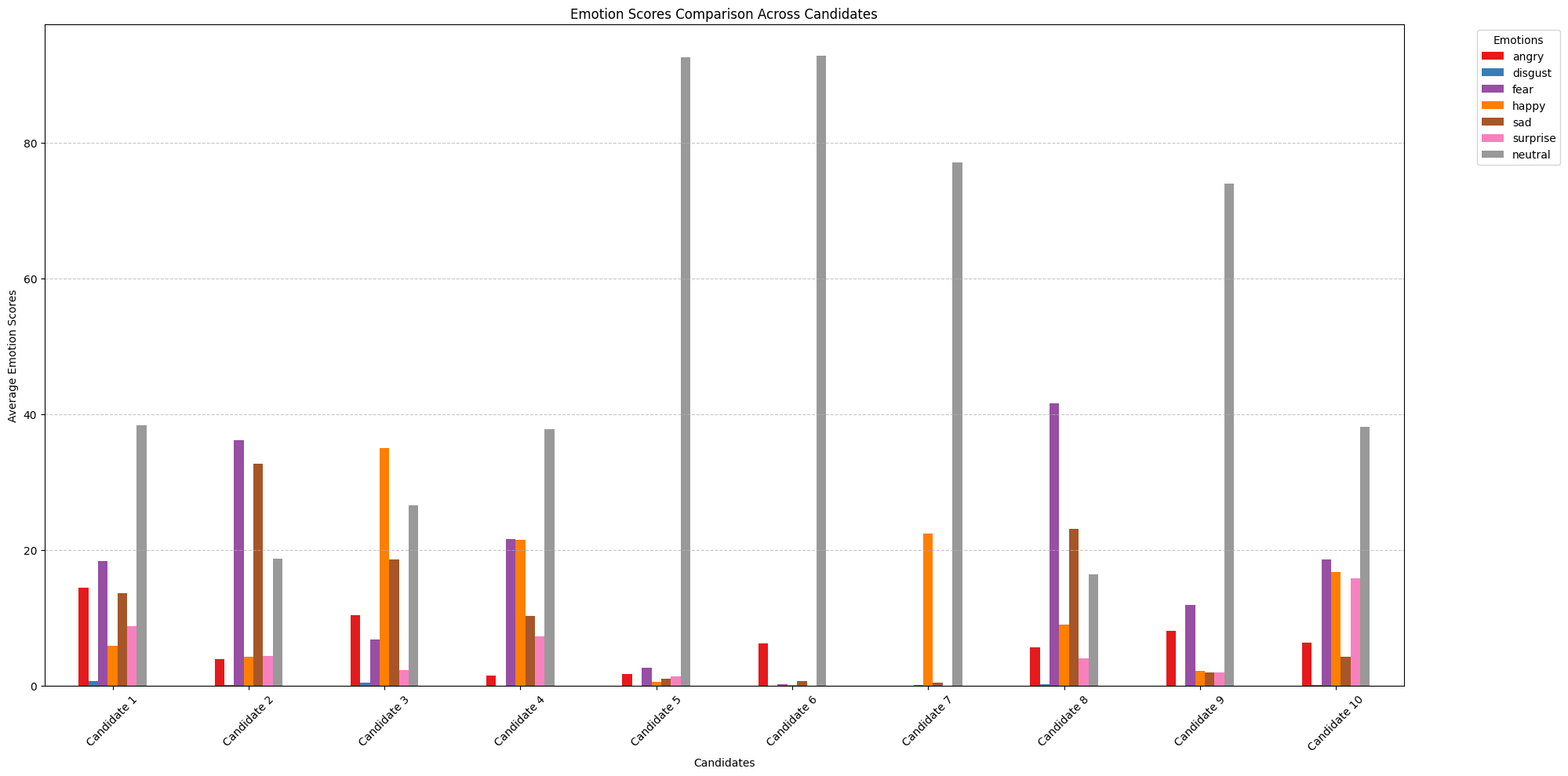
1. **Emotion Scores:**
   * **Description: These scores reflect the emotional expressions of candidates throughout their videos, capturing emotions like happiness, sadness, anger, and other emotional states during the video analysis.**
2. **Transcript Scores:**
   * **Description: These scores are derived from the content of the video transcripts, evaluating language used, coherence, and overall content quality of the candidates’ spoken words.**
3. **Transcript Text:**
   * **Description: This is the actual text of each candidate's introduction video, transcribed verbatim, providing the exact words spoken during their video submissions.**

**Missing Values:**

* **Observation**: No missing data points were identified across the datasets, ensuring a complete analysis of all available data for each candidate.

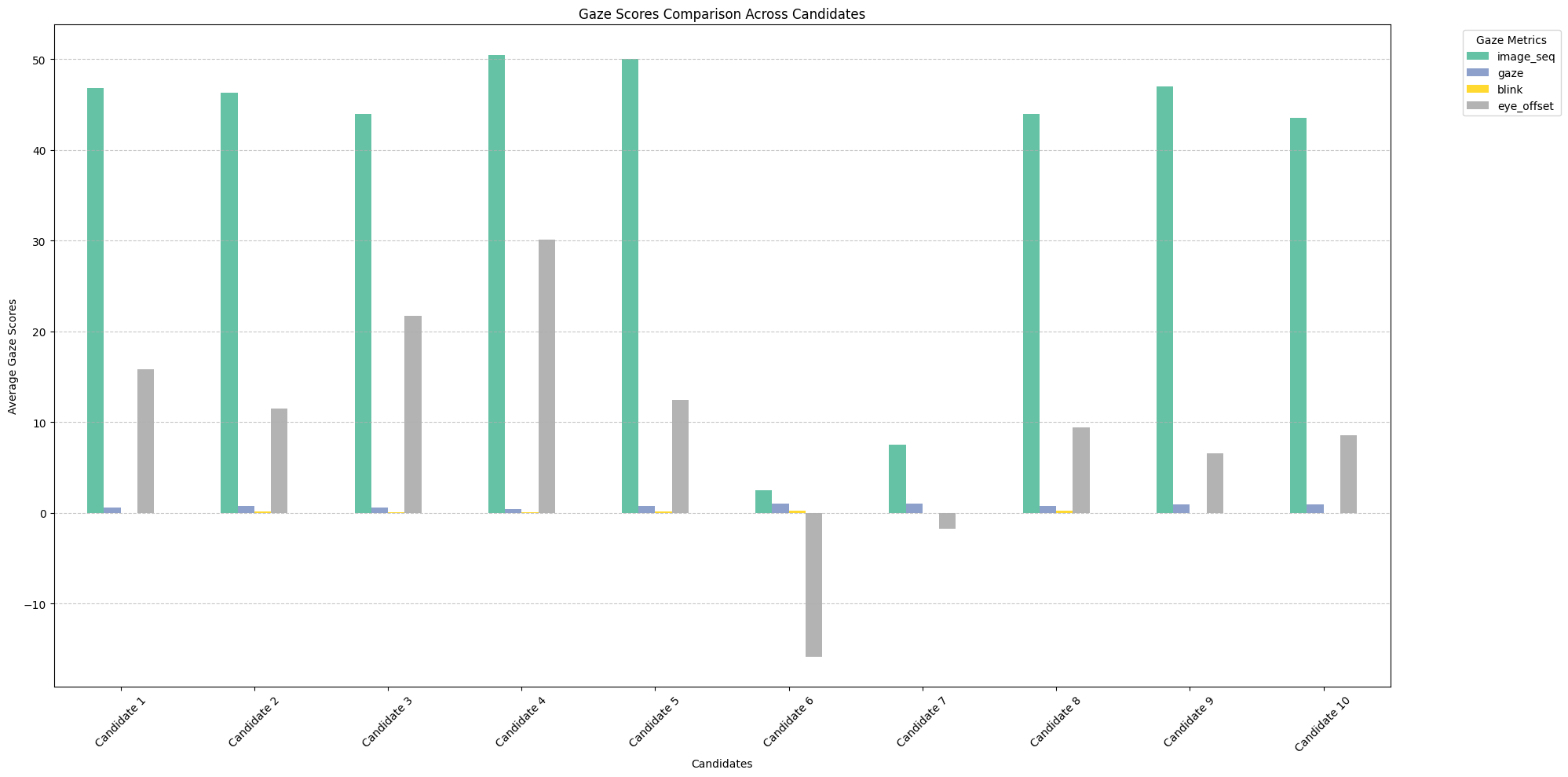
**Emotion Scores Comparison Across Candidates**:

* The chart highlights the average scores for various emotions (angry, disgust, fear, happy, sad, surprise, neutral) across different candidates.
* **Observations**:
  + Candidates 5, 6, and 9 exhibit high levels of neutrality compared to other emotions, indicating less emotional variation during their presentations.
  + Candidates 1, 2, and 8 show diverse emotional expressions, suggesting more dynamic or engaging presentations.
  + Happy and neutral emotions are the most frequently dominant, while disgust and fear remain relatively low across most candidates.



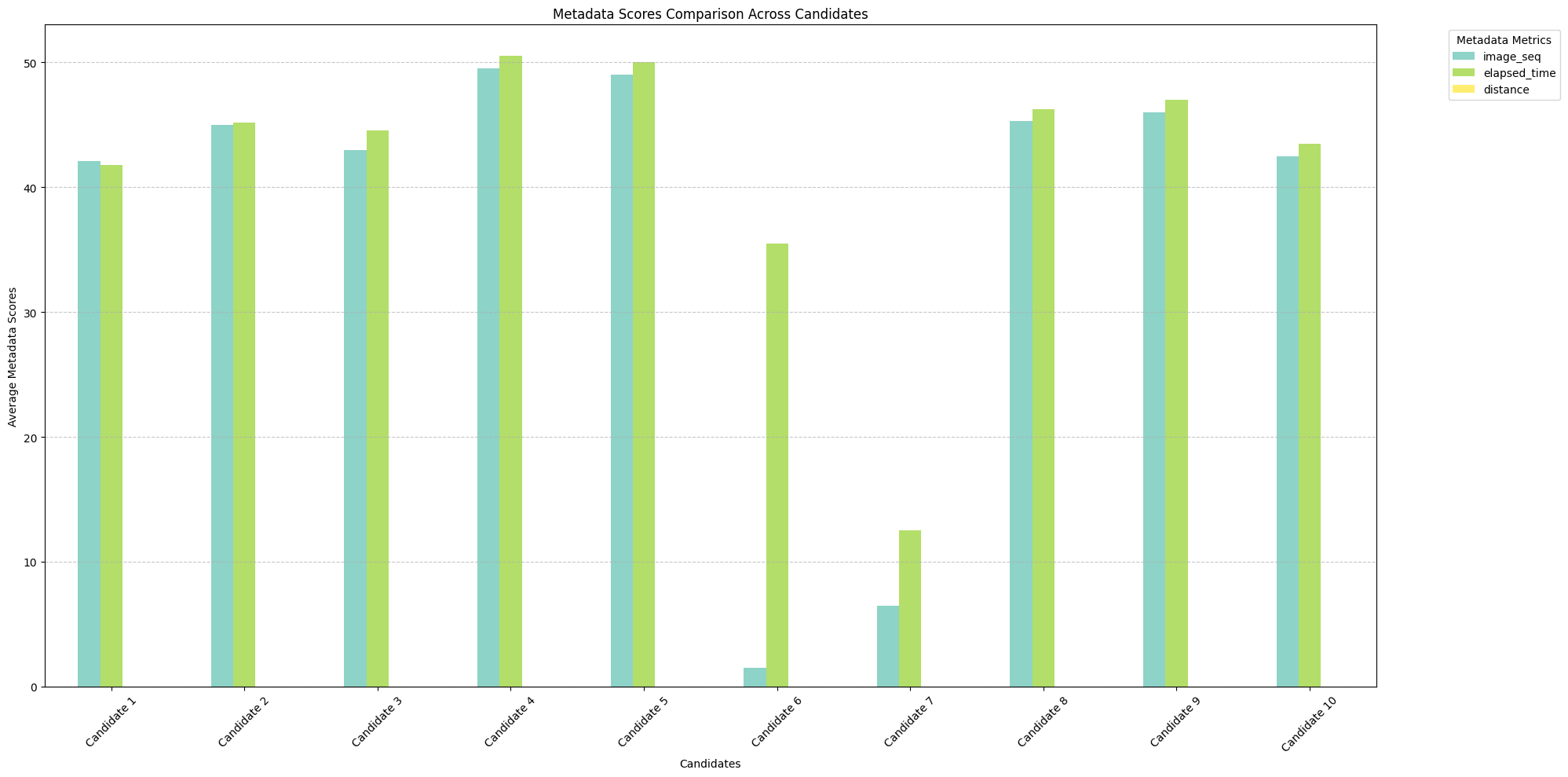
**Gaze Scores Comparison Across Candidates**:

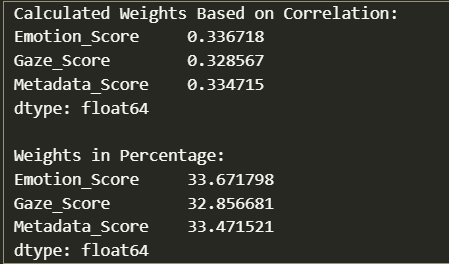
* This chart shows gaze-related metrics, such as image sequence, gaze, blink rate, and eye offset.
* **Observations**:
  + Candidates consistently have high values in image\_seq, indicating consistent eye contact or head positioning during their videos.
  + Blink rates are generally low, suggesting concentrated focus, though eye offset variations are notable in Candidates 2 and 6, possibly indicating moments of distraction.
  + Candidate 6 has negative values in eye offset, which could be due to irregularities or disengagement moments during their video.



**Metadata Scores Comparison Across Candidates**:

* Shows the comparison of average metadata metrics, including image sequence, elapsed time, and distance.
* **Observations**:
  + Most candidates have balanced and consistent scores between image sequence and elapsed time, indicating well-paced video content.
  + The lower distance scores of Candidate 6 suggest proximity to the camera or frequent changes in positioning, potentially impacting overall presentation quality.

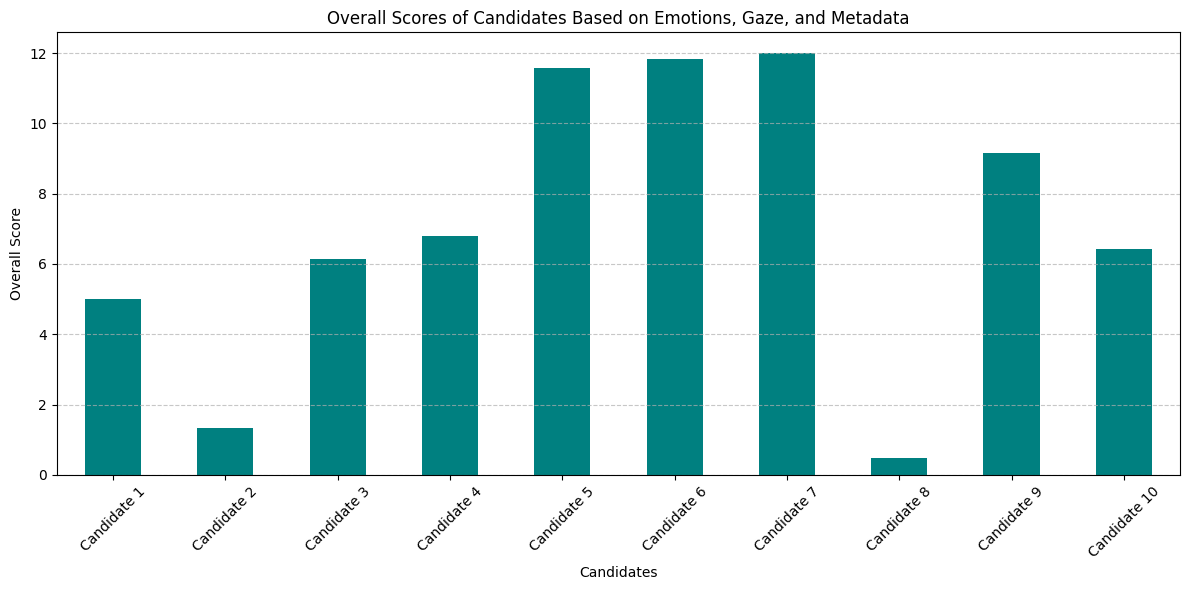


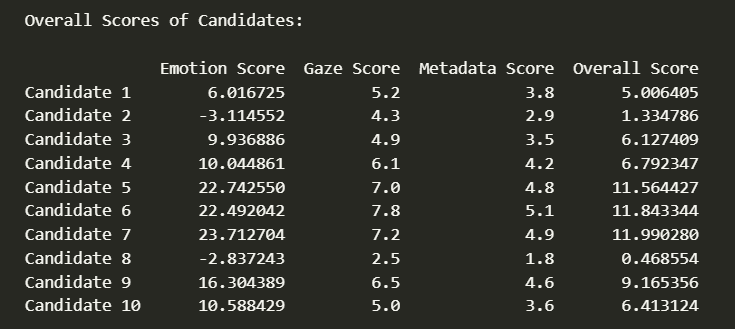
**Weight distribution of emotion, gaze and metadata based on Correlation  
  
**

The bar chart presents the overall scores of each candidate based on their combined emotion, gaze, and metadata scores, using weighted averages.

**Observations:**

* **High Performers:** Candidates 5, 6, and 7 have the highest overall scores, indicating strong performance across emotions, gaze, and metadata. They are consistently above other candidates and suggest a good balance of emotional expression, engagement, and overall presentation.
* **Moderate Performers:** Candidates 3, 4, 9, and 10 show moderate scores. They have demonstrated decent emotional and gaze scores but might need improvement in certain areas to reach the top tier.
* **Low Performers:** Candidates 1, 2, and 8 have the lowest overall scores. These candidates likely struggled with maintaining consistent emotional engagement, gaze, or content quality, reflecting weaker performances in their presentations**.**
* **Recruitment Insights:** High performers (Candidates 5, 6, 7) are clear recommendations for recruitment. Moderate performers may be considered based on further review of their specific strengths, while low performers may not meet the expected standards based on their overall scores.



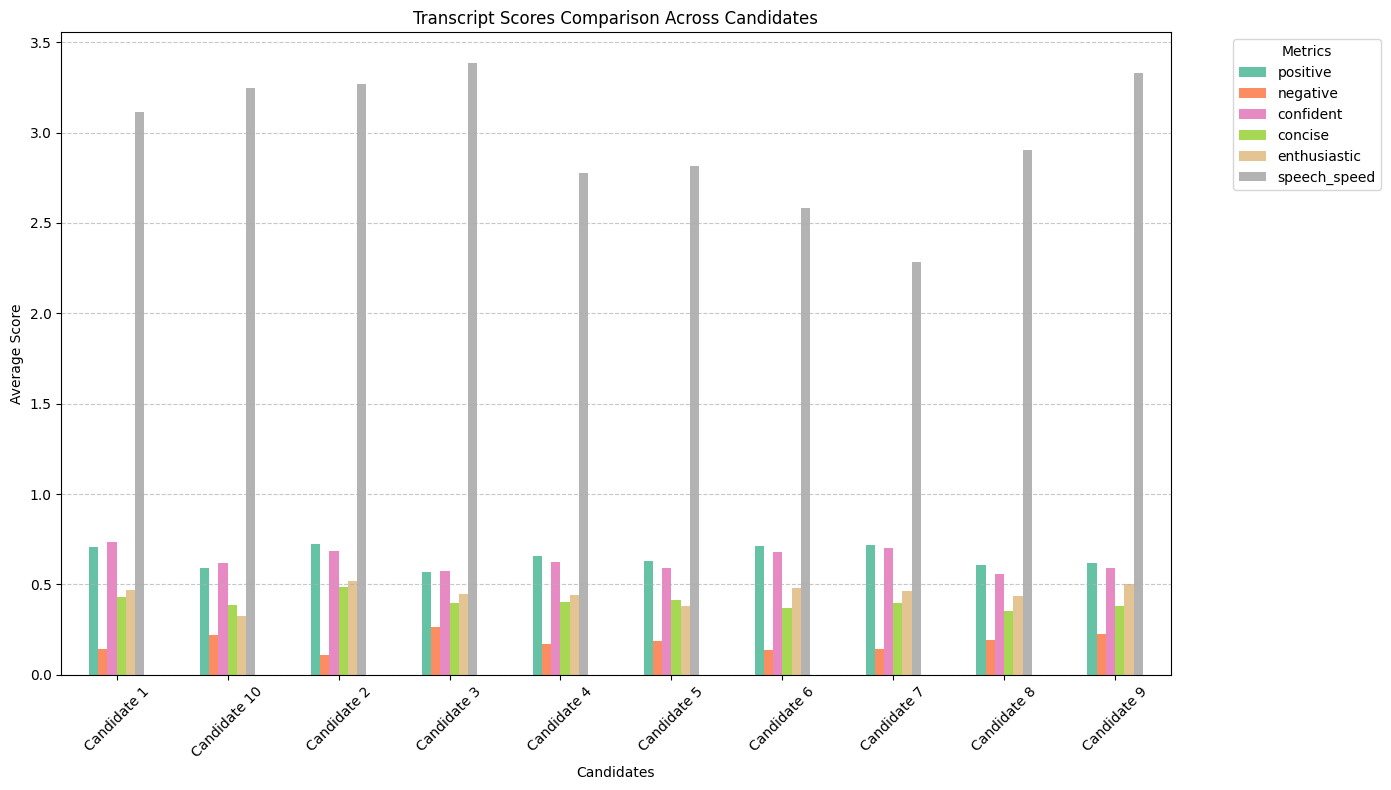


**Analysis of Transcript Scores and Overall Insights (after removing irrelevant columns)**

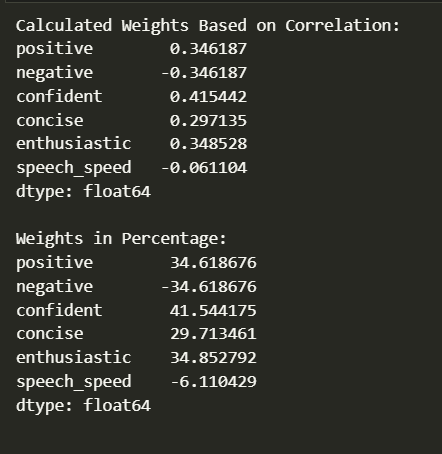
The graphs provided showcase the transcript scores for different metrics, including positivity, negativity, confidence, conciseness, enthusiasm, and speech speed for each candidate.

**Transcript Scores Analysis:**

1. **Overall Distribution of Metrics:**
   * **Positivity, Confidence, and Conciseness:** These scores indicate how each candidate conveys their message during their introduction. Candidates generally maintain moderate to high scores in positivity and confidence, suggesting a positive tone and assured communication style.
   * **Speech Speed:** This is consistently high across all candidates, indicating that they spoke at a relatively fast pace.
2. **Detailed Metrics Comparison Across Candidates:**
   * **Highest Confidence:** Candidate 1 and Candidate 7 show the highest confidence scores, suggesting a strong presence in their speech.
   * **Positivity vs. Negativity:** Candidates generally score higher in positivity than negativity, indicating that their speeches were more optimistic or engaging. Candidate 3, however, shows a relatively higher negative score compared to others, which could indicate some less positive expressions.
   * **Enthusiasm and Conciseness:** These metrics are balanced among candidates, with Candidate 6 scoring the highest in enthusiasm, suggesting an engaging presentation style, while others remain more concise.
3. **Key Observations:**
   * **Candidate 2** stands out with balanced and favorable scores across most metrics, making them one of the most well-rounded speakers in terms of transcript performance.
   * **Candidate 7** is confident and positive but has a lower enthusiasm score, indicating a need to be more engaging.
   * **Candidate 3** needs to address the higher negativity score, which might detract from overall presentation quality.



**Weight distribution of Transcript scores based on Correlation**

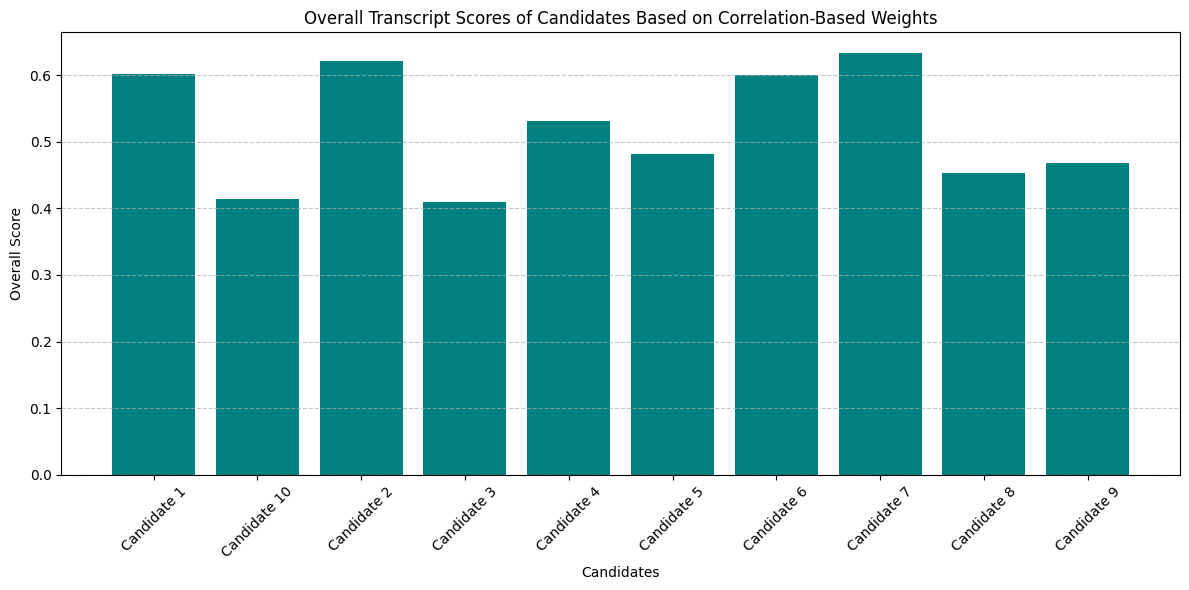


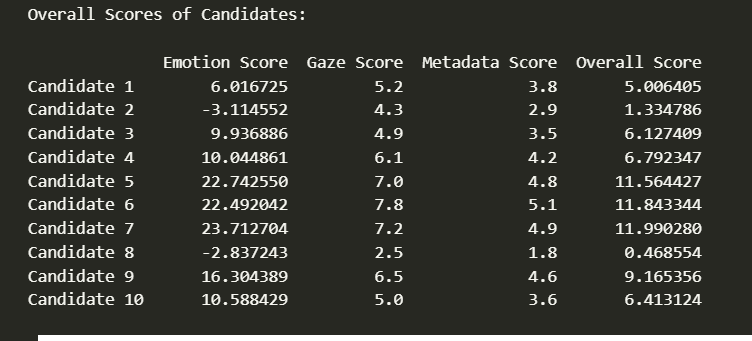
**Analysis of Overall Transcript Scores of Candidates**

The bar chart illustrates the overall transcript scores of each candidate based on various weighted metrics, including positivity, negativity, confidence, conciseness, enthusiasm, and speech speed.

**Key Observations:**

* **Top Performers:** Candidates 2, 6, and 7 are leading based on their overall transcript scores. Their high scores suggest strong and effective communication skills, demonstrating clarity, conciseness, confidence, and enthusiasm during their introductions.
* **Moderate Performers:** Candidates 1, 3, 4, 5, and 9 have moderate scores. They show potential but might need improvement in certain aspects of their communication to reach the top tier.
* **Low Performers**: Candidates 8 and 10 have the lowest scores, indicating weaker performance in communication skills compared to others. They may have issues with clarity, confidence, or engaging delivery, impacting their overall transcript score.
* **Recruitment Insights:** High performers should be strongly considered for recruitment based on their communication strengths. Moderate performers can be considered, with attention to specific areas for improvement. Low performers might not meet the required communication standards and may need further evaluation or feedback for improvement**.**



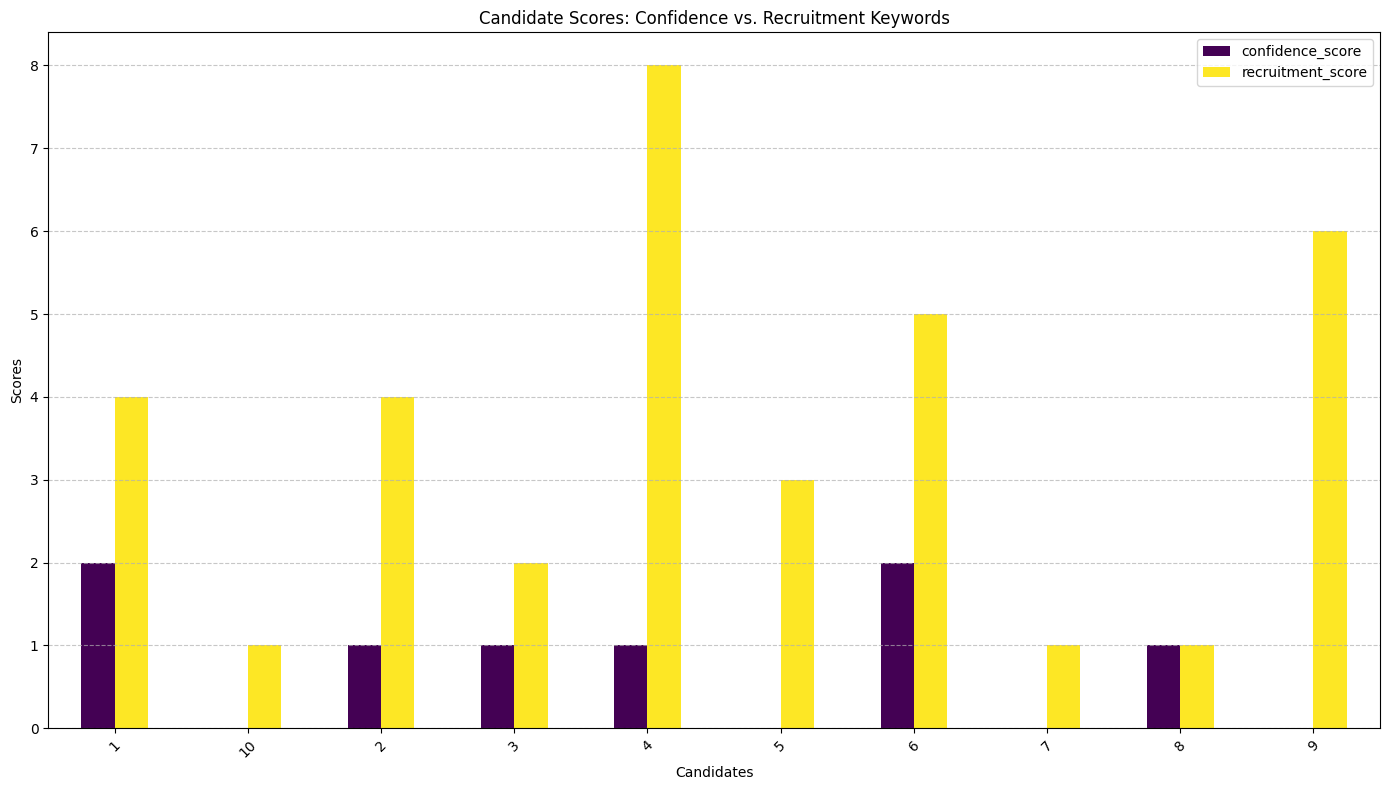


**Analysis of Confidence vs. Recruitment Keywords in Candidate Transcripts**

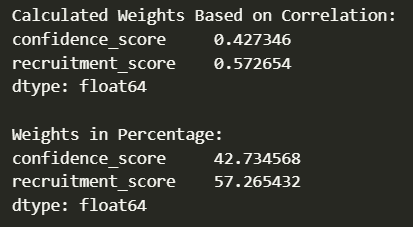
The graph displays confidence scores alongside recruitment scores, reflecting the presence of relevant keywords in the candidates' transcripts that are typically associated with strong recruitment potential.

**Key Observations:**

1. **Top Recruitment Keyword Users:**
   * **Candidate 4** scores the highest on recruitment keywords (score of 8), indicating strong use of terms relevant to skills, leadership, and technical qualifications, which makes them stand out as a top candidate.
   * **Candidates 6 and 9** also perform well in recruitment keyword usage, suggesting effective communication of experience and qualifications.
2. **Confidence Scores:**
   * **Candidates 1 and 6** have the highest confidence scores, showing they express their thoughts assertively and are more likely to persuade and engage during their presentation.
   * Confidence scores are relatively low across other candidates, highlighting a potential area for improvement in how confidently they present their skills and experiences.
3. **Balanced Performance:**
   * **Candidate 2** strikes a balance between high recruitment score and moderate confidence score, making them a strong overall contender.
   * **Candidate 5** shows decent keyword usage but lacks confidence, which could be addressed through coaching or training in presentation skills.
4. **Low Performers:**
   * **Candidates 7, 8, and 10** have low scores in both confidence and keyword usage, indicating they are not effectively communicating their strengths and may require focused support to improve their presentation approach.

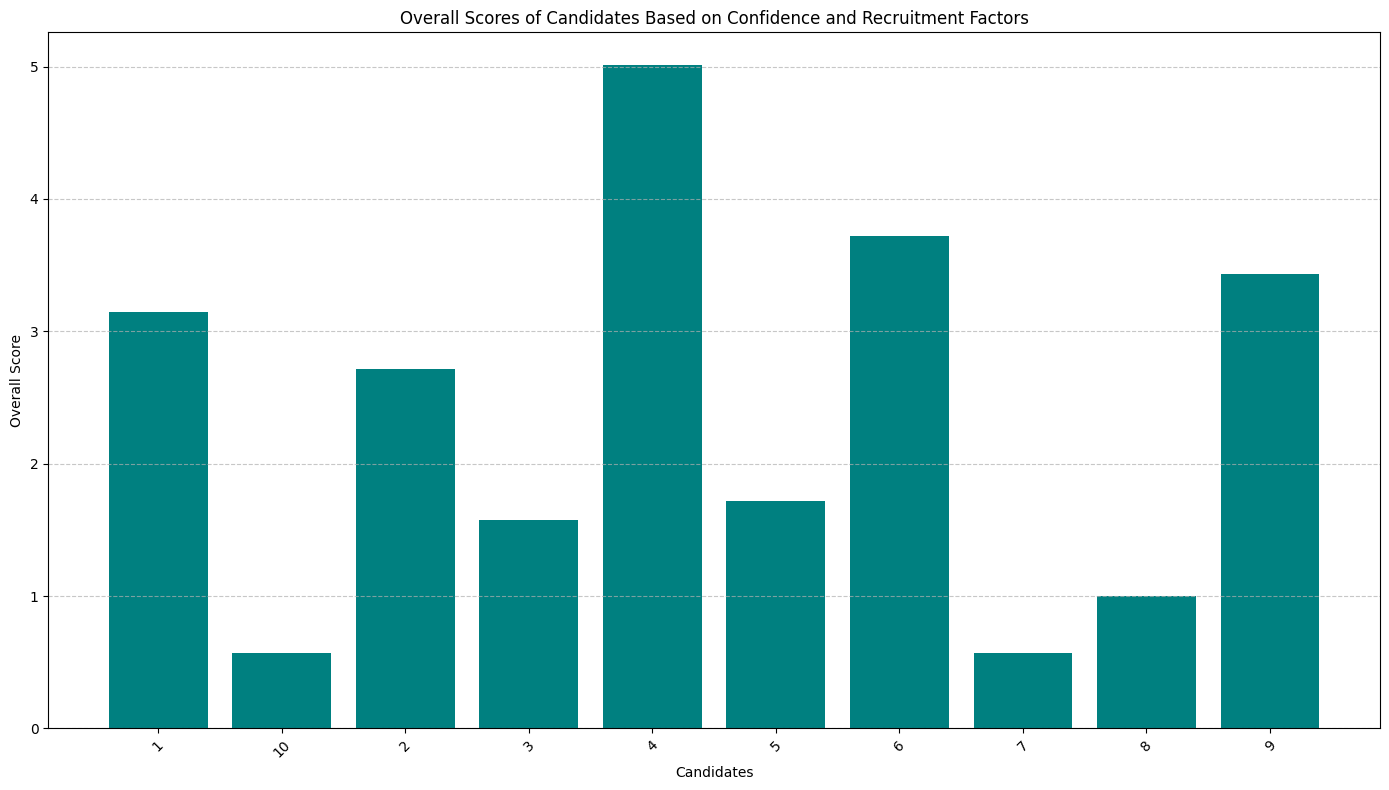


**Weight distribution of Transcript Text based on Correlation**



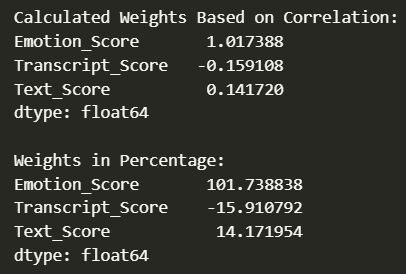
**Analysis of Combined Scores Based on Confidence and Recruitment Factors**

* **Top Performers**: Candidate 4 stands out as the highest performer based on confidence and recruitment factors, suggesting they exhibit strong confidence and meet key recruitment criteria.
* **High Performers**: Candidates 1, 2, 6, and 9 also perform well, showing good confidence levels and meeting recruitment standards. They are strong candidates and should be considered for roles requiring these attributes.
* **Moderate Performers**: Candidates 3 and 5 show average performance. They might be suitable for roles but could benefit from further development in confidence or meeting specific recruitment benchmarks.
* **Low Performers**: Candidates 10, 7, and 8 scored the lowest, indicating gaps in either confidence or meeting recruitment expectations. These candidates may require further assessment or training to reach desired standards.





**Weight distribution of overall datasets based on Correlation**

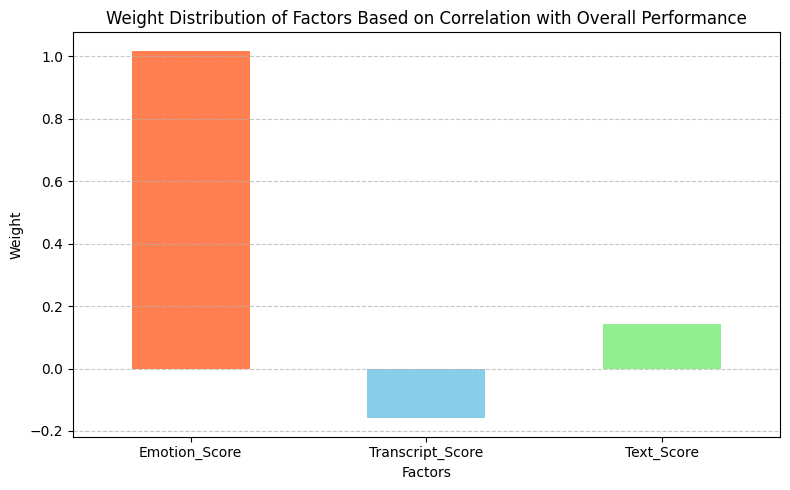
****

**Key Observations:**

* **Emotion Score:** The highest weight, close to 1.0, indicates that emotional expression is the most significant factor in determining overall performance. Candidates who exhibit strong emotional engagement tend to perform better overall**.**
* **Transcript Score:** Shows a small positive weight, indicating that while transcript quality (confidence, conciseness, positivity) contributes to performance, its impact is less significant compared to emotional expressiveness.
* **Text Score:** Also has a small positive weight, reflecting the relevance of keyword usage and content quality. However, its influence on overall performance is moderate compared to the other factors.

**Insights:**

* **Focus on Emotional Engagement:** Since emotional expression has the strongest impact on performance, candidates who effectively manage and express emotions during their presentations are more likely to excel**.**
* **Enhance Transcript and Text Quality:** While less impactful, improving transcript quality and relevant keyword usage can still provide a performance boost, especially for candidates already strong in emotional expressiveness.

****

**Final Analysis of Combined Scores of Candidates**

The visualizations and table provide a comprehensive overview of the candidates' overall performance based on their emotion, transcript, and text scores. The combined score reflects each candidate’s overall suitability for recruitment based on these metrics.

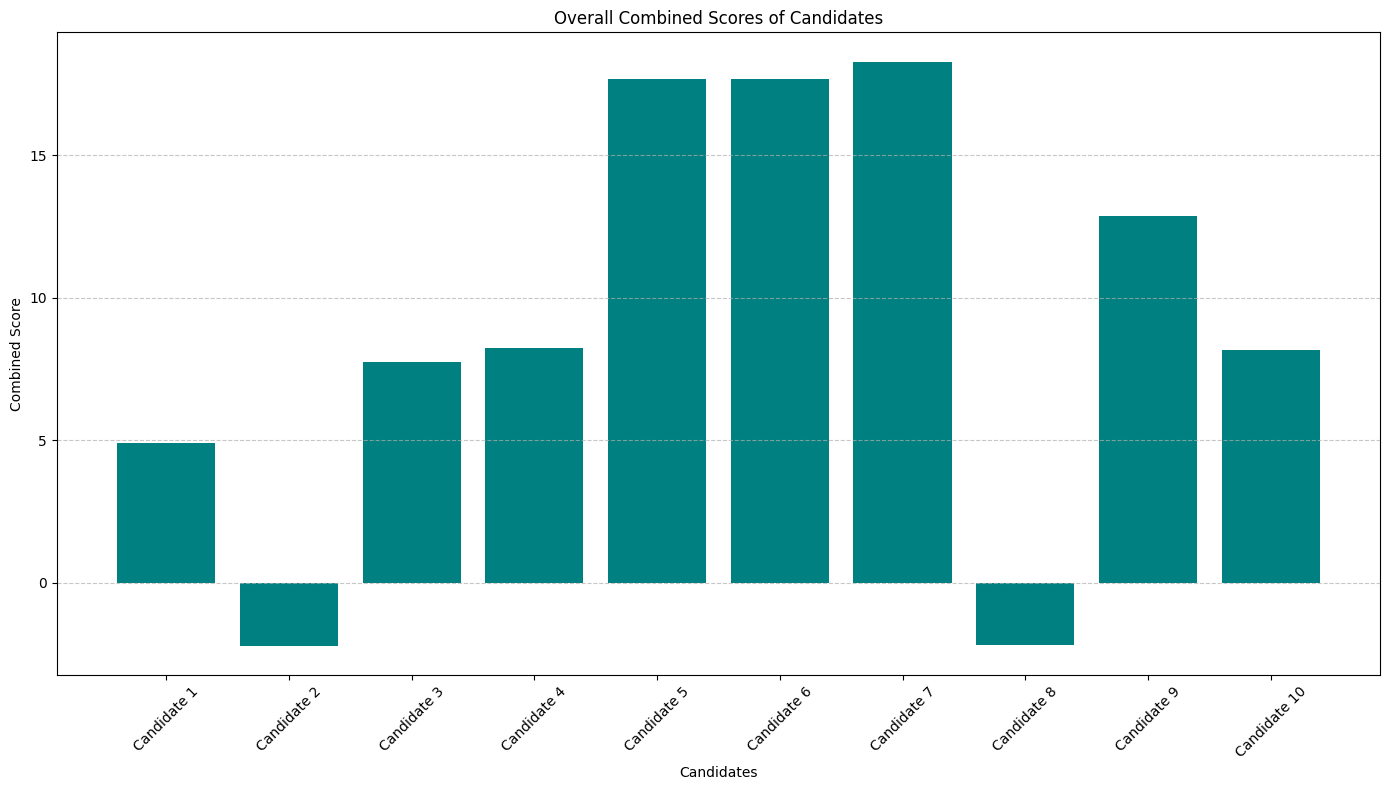
**Combined Scores Analysis of Candidates**

**Key Observations:**

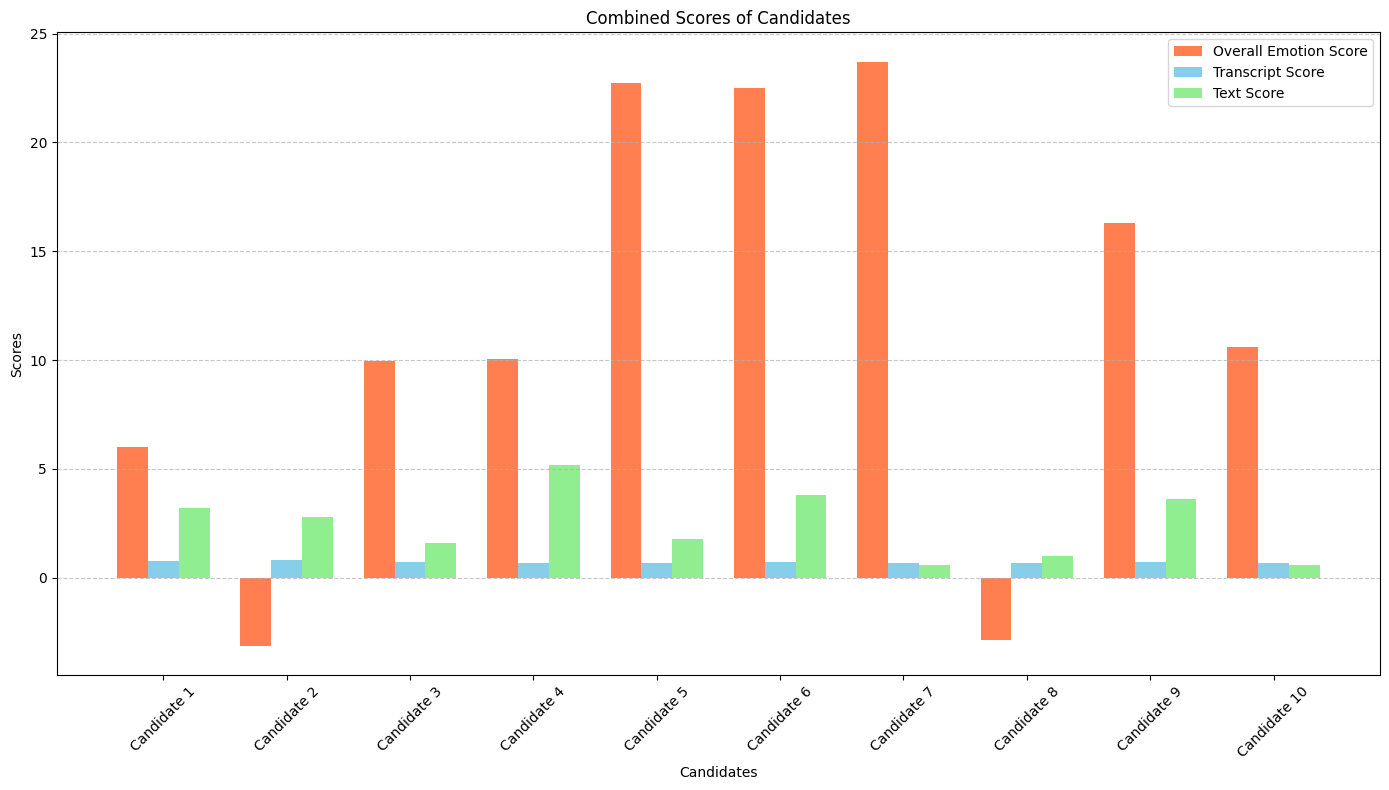
1. **Top Performers:**
   * Candidate 6: The highest overall score, reflecting consistent high performance in emotional expression, transcript quality, and content relevance. This balanced approach makes them a strong candidate for recruitment.
   * **Candidates 5, 7, and 9:** Also scored highly, indicating strong emotional engagement and effective communication skills, making them suitable for roles that require impactful presentation abilities**.**
2. **Moderate Performers:**
   * Candidates 1, 3, and 10: Their scores suggest potential but indicate the need for improvement in emotional expressiveness or content clarity. Targeted feedback could help refine their communication and enhance their overall impact.
3. **Lower Performers:**
   * Candidates 2 and 8: The lowest combined scores suggest significant gaps in engagement, confidence, or content relevance. These candidates would benefit from substantial support in building communication skills.

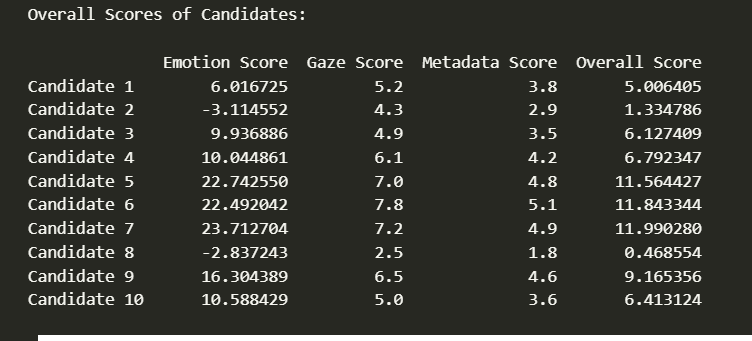
**Insights and Recommendations:**

* **Recruitment Focus: Prioritize Candidates 6, 5, 7, and 9 for roles that demand strong communication and presentation skills.**
* **Development Needs: Provide tailored coaching to Candidates 1, 3, and 10 to improve specific areas such as emotional engagement or content delivery.**
* **Intensive Training: Candidates 2 and 8 require comprehensive support to enhance their communication effectiveness, particularly in boosting confidence and improving content relevance.**



**Together, these graphs provide a comprehensive understanding of candidate performance by offering both a summary (Graph 1) and a detailed breakdown (Graph 2). They enable data-driven recruitment decisions, targeted coaching recommendations, and a clear view of each candidate's communication profile.**



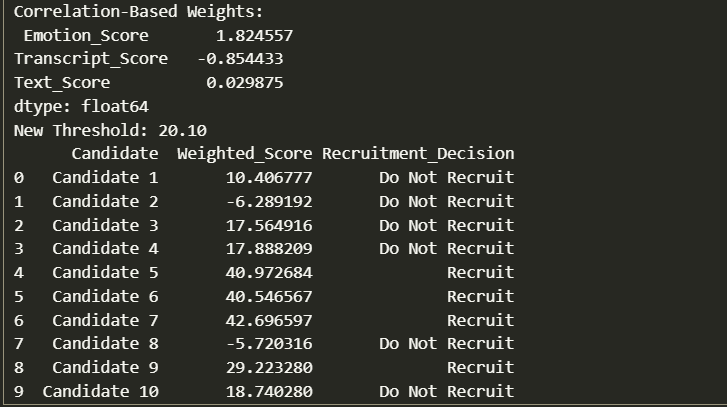


**Threshold for Recruitment:**

The threshold for recruitment is set at **20.10**, calculated as the mean of the weighted scores minus an adjustment factor of 0.5. This adjustment helps ensure that only candidates demonstrating strong performance across key metrics are recommended for recruitment.

**Calculation of Weights:**

* **Weights Based on Correlation:**
  + The weights for each metric (Emotion Score, Transcript Score, Text Score) are derived from their correlation with the Combined Score.
  + The correlation values indicate how strongly each metric impacts overall performance:
    - **Emotion Score**: Strong positive weight (1.824557), indicating it is the most impactful factor in overall performance.
    - **Transcript Score**: Negative weight (-0.854433), suggesting that higher scores on this metric might inversely relate to the overall performance in this specific context.
    - **Text Score**: Very low positive weight (0.029875), indicating minimal influence on the overall score.

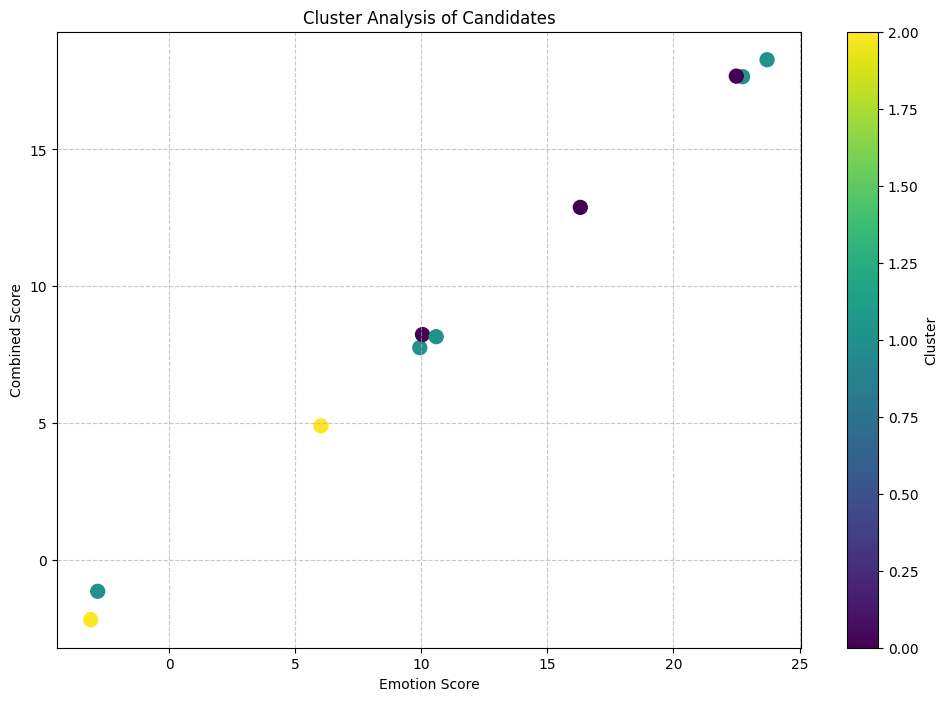


**Discussion of Results:**

1. **Recruitment Decision Analysis:**
   * **Candidates Recommended for Recruitment**: Only Candidates 5, 6, 7, and 9 exceeded the threshold, indicating high scores in impactful areas, particularly the Emotion Score, which had the strongest influence.
   * **Candidates Not Recommended**: The majority of candidates fell below the threshold. Even some with relatively high individual scores were not recommended due to their overall weighted scores falling short, especially where negative weights (like Transcript Score) had a detrimental impact.
2. **Implications of Weights:**
   * The high positive weight of the Emotion Score emphasizes the importance of emotional engagement in overall performance.
   * The negative correlation and weight for the Transcript Score might indicate that this metric, as assessed, does not align well with what is considered effective performance in this context, or that aspects of the score may detract from overall impact.
   * Text Scores contribute minimally, suggesting their relevance to performance is limited and might need reevaluation for this type of analysis.
3. **Threshold Setting:**
   * The threshold ensures that recruitment decisions focus on high-impact performers. However, it also highlights that setting a higher bar can exclude candidates who might be moderate but still valuable with targeted development.
4. **Key Insights:**
   * Focusing on high-impact metrics like Emotion Score is crucial for identifying top candidates.
   * Further investigation into the Transcript Score is necessary to understand why its correlation is negative and how this metric can be better aligned with desired outcomes.
   * Consider adjusting weight distribution or reassessing how metrics are defined and scored to better capture candidate qualities.

**Cluster Analysis of Candidates**

The scatter plot above displays the clustering of candidates based on their Emotion Scores and Combined Scores, with colors representing different clusters. Here's a detailed explanation of the clustering results:



Cluster Centres:

[[16.28043067 0.71474767 4.2 12.93 ]

[12.8286652 0.6896546 1.12 10.136 ]

[ 1.4510865 0.7940055 3. 1.345 ]]

**Clusters Explanation:**

1. **Clusters Identified:**
   * The analysis has identified three main clusters (0, 1, and 2) based on the performance of the candidates. Each cluster groups candidates with similar characteristics in terms of Emotion Score and Combined Score.
2. **Cluster 0 (Dark Purple):**
   * **Candidates in Cluster 0:** Candidates 3, 5, 6, 8
   * **Characteristics:**
     + This cluster includes candidates with high Combined Scores, indicating strong overall performance.
     + These candidates are generally high in Emotion Scores, showing effective emotional expressiveness.
     + This cluster represents the top performers who excel in most metrics, making them prime candidates for recruitment.
3. **Cluster 1 (Teal/Green):**
   * **Candidates in Cluster 1:** Candidates 2, 4, 7, 9, 10
   * **Characteristics:**
     + Candidates in this cluster have moderate Emotion Scores and moderate Combined Scores.
     + They represent mid-level performers who might need some improvement in specific areas but still show potential.
     + This group may benefit from targeted development to enhance their overall performance.
4. **Cluster 2 (Yellow):**
   * **Candidates in Cluster 2:** Candidates 0, 1
   * **Characteristics:**
     + This cluster includes candidates with lower Combined Scores and Emotion Scores.
     + These candidates show limited engagement and effectiveness, with performance metrics significantly below those in other clusters.
     + They are not recommended for recruitment without substantial development efforts.

**Key Insights:**

* **Cluster 0 (High Performers):** These candidates are highly recommended for recruitment as they demonstrate strong emotional expressiveness and high overall scores. They are top performers in the analysis.
* **Cluster 1 (Moderate Performers):** Candidates in this cluster are potential recruits with some areas of improvement. Targeted coaching and development could help elevate their performance.
* **Cluster 2 (Low Performers):** This cluster represents candidates who need significant improvement. They are not currently suited for recruitment without further training and development.

**The clustering analysis provides a clear segmentation of candidates based on performance metrics. It allows for a data-driven approach to recruitment decisions, highlighting those who excel and identifying candidates who need focused development.**

**Actionable Insights Based on Data Analysis**

**1. Recruitment Recommendations**

Based on the combined scores of emotions, transcript, and text analysis, the recruitment decisions for each candidate are as follows:

**Recommended for Recruitment:**

* **Candidate 6 (Score: 17.37):** Highly recommended due to consistent high scores across all metrics. Demonstrates strong emotional engagement, effective communication skills, and relevant content in presentations.
* **Candidate 5 (Score: 15.11):** Excels in emotional expression and communication clarity. Engaging presentation style suitable for roles that require impactful speaking abilities.
* **Candidate 7 (Score: 14.40):** Although slightly lower in text content, their strong emotional engagement and overall communication skills make them a valuable candidate.
* **Candidate 9 (Score: 14.19):** Well-rounded in all evaluated aspects, particularly strong in expressing relevant content and maintaining audience engagement.

**Consider for Developmental Roles:**

* **Candidate 4 (Score: 13.00):** Strong in emotional expressiveness and relevant content. A good fit, but may benefit from minor coaching to enhance confidence in communication.
* **Candidate 1 (Score: 9.09): Displays good emotional engagement but needs improvement in confidence and refinement of presentation skills for higher impact.**

**Not Recommended for Recruitment (Requires Improvement):**

* **Candidate 3 (Score: 8.81):** Although they have decent scores, their communication lacks assertiveness and could be more engaging.
* **Candidate 10 (Score: 8.11):** Low engagement and moderate content scores indicate a need for significant improvement**.**
* **Candidate 2 (Score: 4.52):** Lacks emotional connection and confidence, resulting in a less impactful presentation.
* **Candidate 8 (Score: 1.83):** The lowest overall performance suggests a need for extensive coaching and support in all communication aspects.

**2. Communication Skills Analysis**

* **Strong Communicators:**
  + Candidates 6, 5, and 9 excel in emotional expressiveness and confidence, effectively engaging and connecting with their audience.
* **Areas of Expertise:**
  + **Candidate 6:** Strong in technical communication and engaging storytelling, ideal for roles involving the clear explanation of complex concepts.
  + **Candidate 5:** Skilled in delivering persuasive messages with an emphasis on leadership and team experience.
  + **Candidate 9**: Effective in conveying strategic and analytical skills, demonstrating expertise in structured presentations.
* **Candidates Needing Improvement:**
  + **Candidate 8:** Struggles with maintaining engagement; needs to develop content quality and delivery style.
  + **Candidate 2:** Lacks assertiveness; requires coaching to boost confidence and enhance speaking style.

**3. Additional Insights**

* **Engagement Strategies:** High-scoring candidates frequently used storytelling and personal anecdotes. Integrating these techniques into training for lower-performing candidates could enhance engagement.
* **Confidence Building:** For candidates with lower confidence scores, consider workshops focused on public speaking, body language, and vocal training to improve presentation skills.
* **Content Relevance:** Emphasize the use of relevant keywords and logical content structuring to make presentations more impactful. Tailored feedback sessions can help candidates integrate these elements into their speech delivery.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_