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.

**2. Data Summary**

**Dataset Overview:**

* **Number of Candidates**: 10
* **Features Analysed**: The analysis focused on three key features:
  + **Emotion Scores**: Evaluated to understand candidates' emotional engagement and expressions during their presentations.
  + **Transcript Scores**: Assessed to gauge the effectiveness and quality of the spoken content in the candidates' transcripts, including metrics like positivity, confidence, and conciseness.
  + **Text Scores**: Measured based on the presence of recruitment-related keywords and the confidence displayed in their delivery, highlighting the candidates' ability to articulate their experience and skills relevant to recruitment.

**Missing Values:**

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

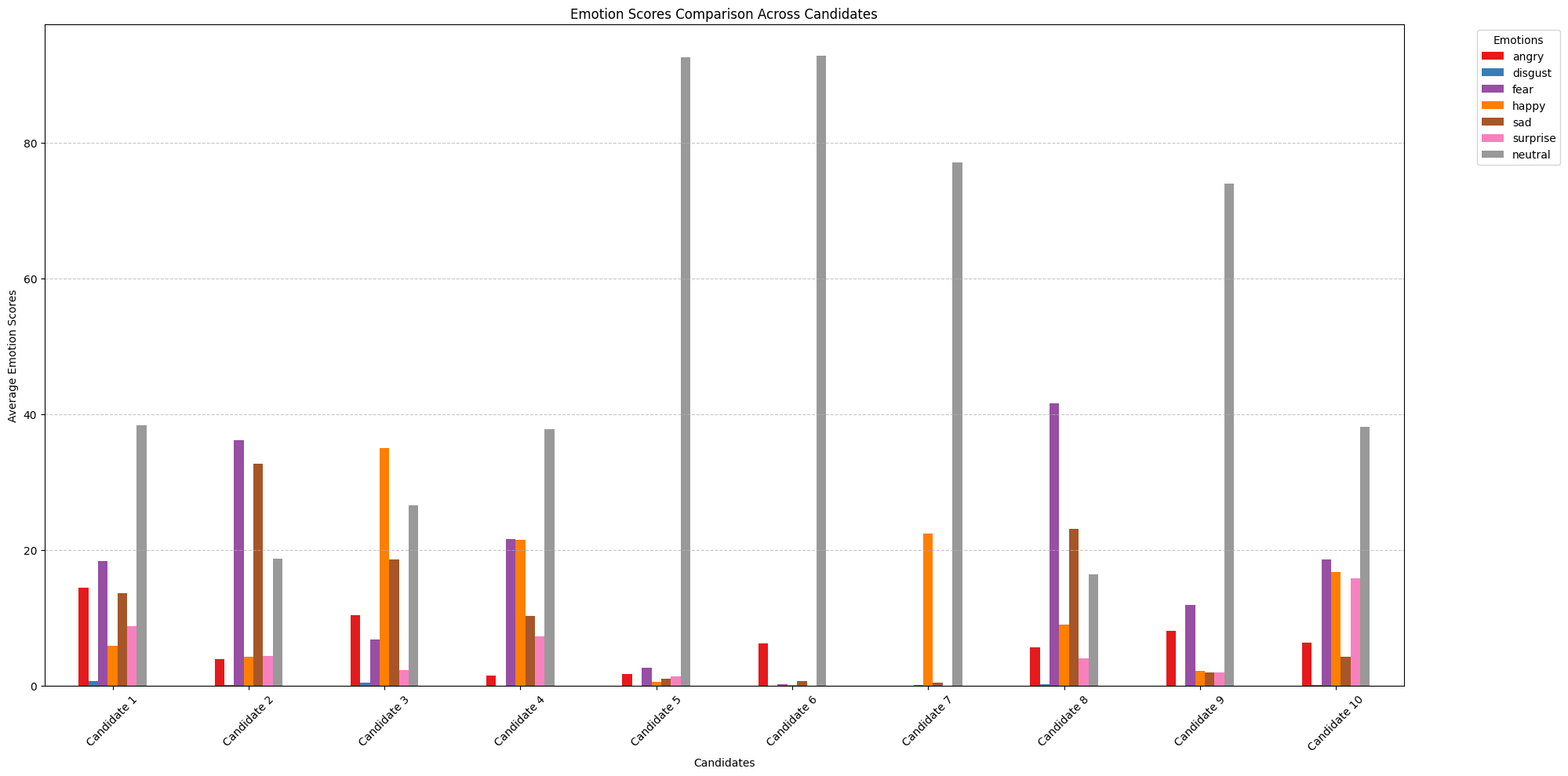
**Descriptive Statistics Summary**

The descriptive statistics are visualized through distribution plots, bar charts, and comparison graphs, highlighting key findings from the EDA:

* **Emotion Scores**: Visuals show variations in emotional expressions like happiness and neutrality. High scores in Candidates 5 and 6 indicate strong engagement, while lower emotional variability in Candidates 2 and 8 suggests a weaker connection.
* **Transcript Scores**: Comparative charts reveal metrics such as confidence and positivity. Top performers like Candidates 2, 6, and 9 demonstrate persuasive speech, while lower conciseness in some suggests room for improvement.
* **Text Scores**: Bar charts highlight the use of impactful keywords. Candidates 4 and 9 excel in using relevant terms, enhancing their presentation quality, while lower-scoring candidates need to integrate more relevant keywords to convey their skills effectively.

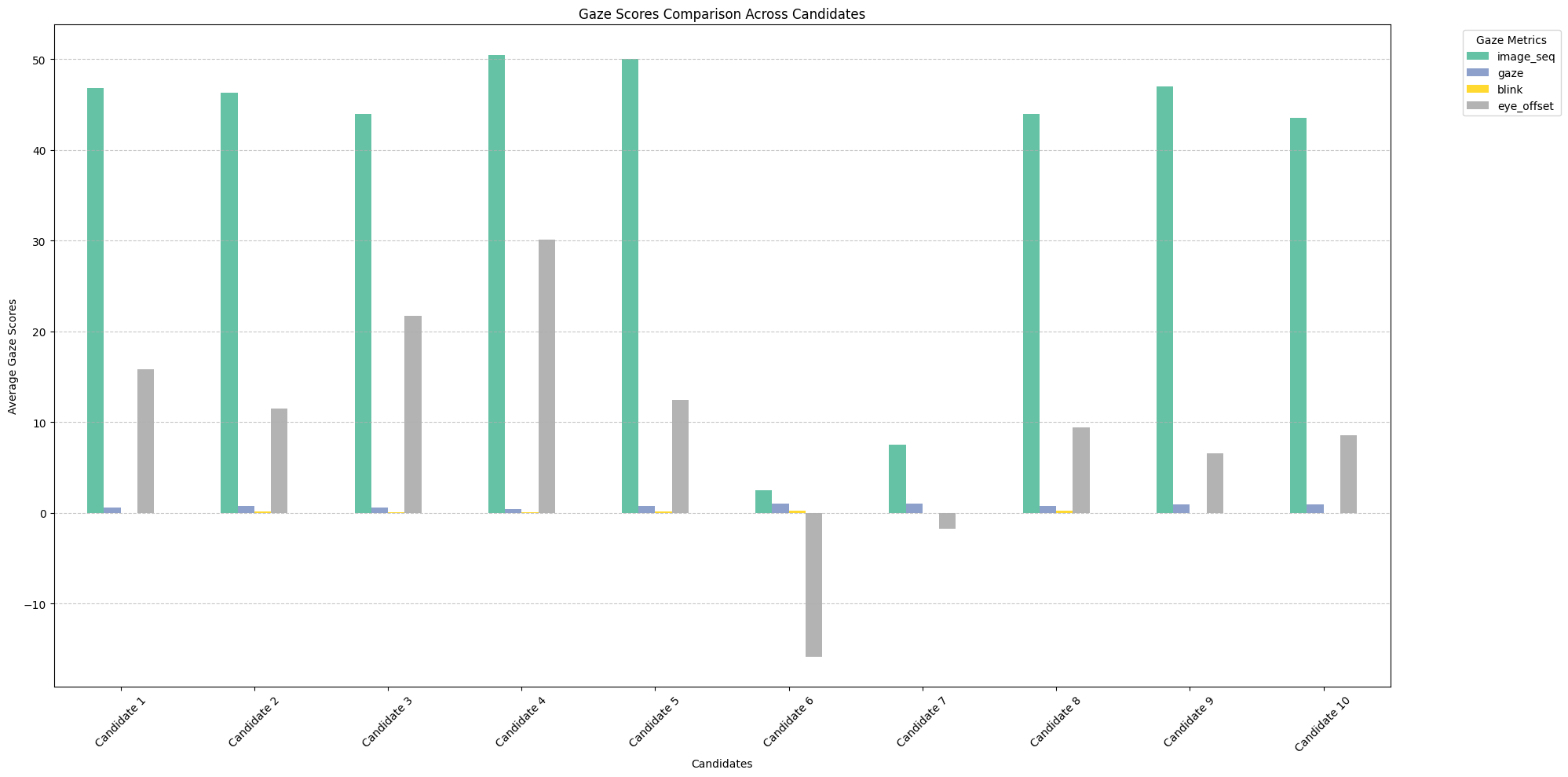
**Emotion Scores Comparison Across Candidates**:

* 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.



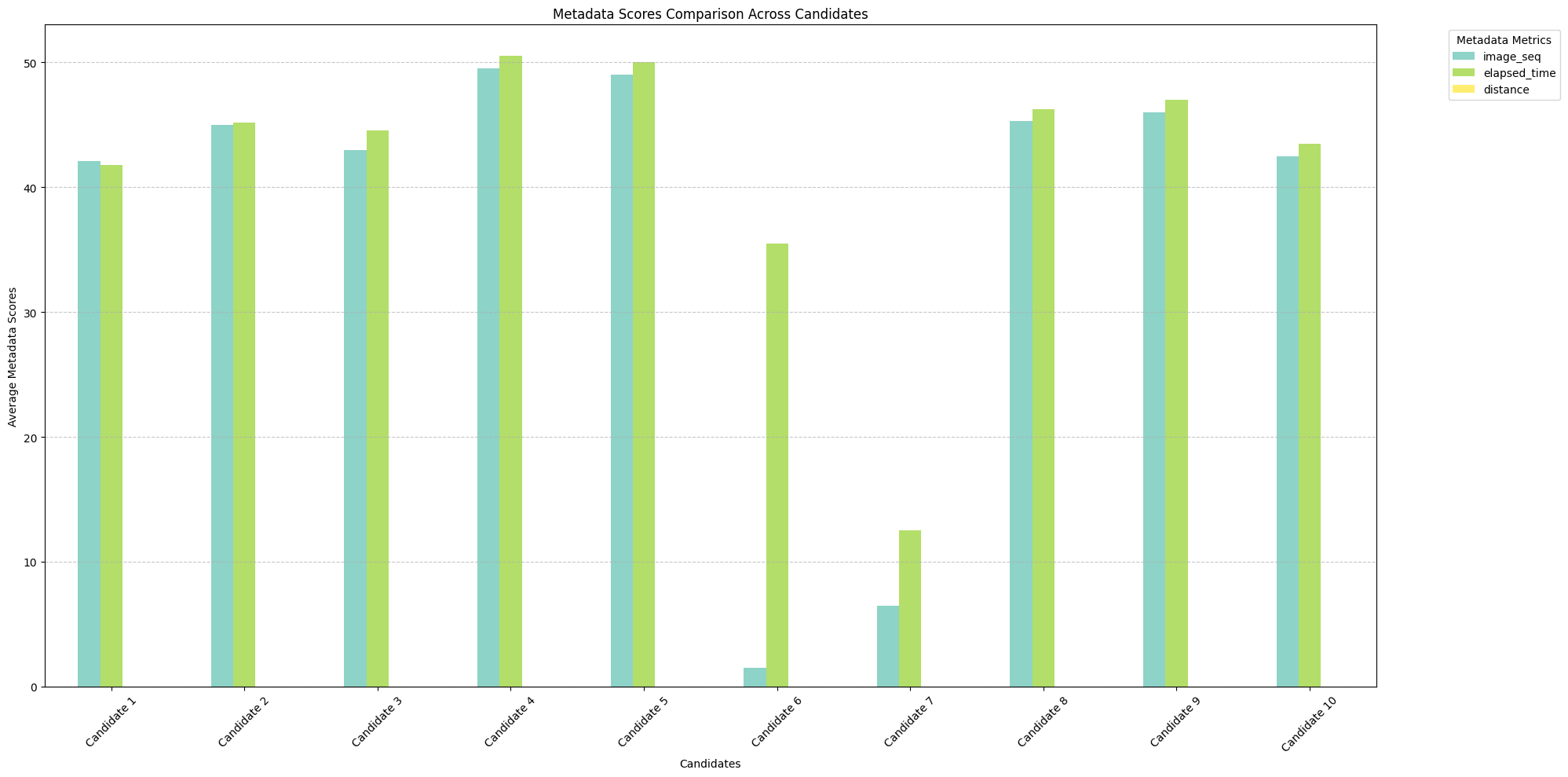
**Gaze Scores Comparison Across Candidates**:

* 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.



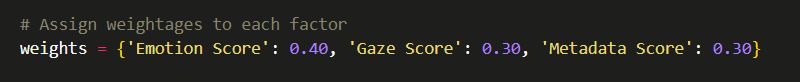
**Metadata Scores Comparison Across Candidates**:

* 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.



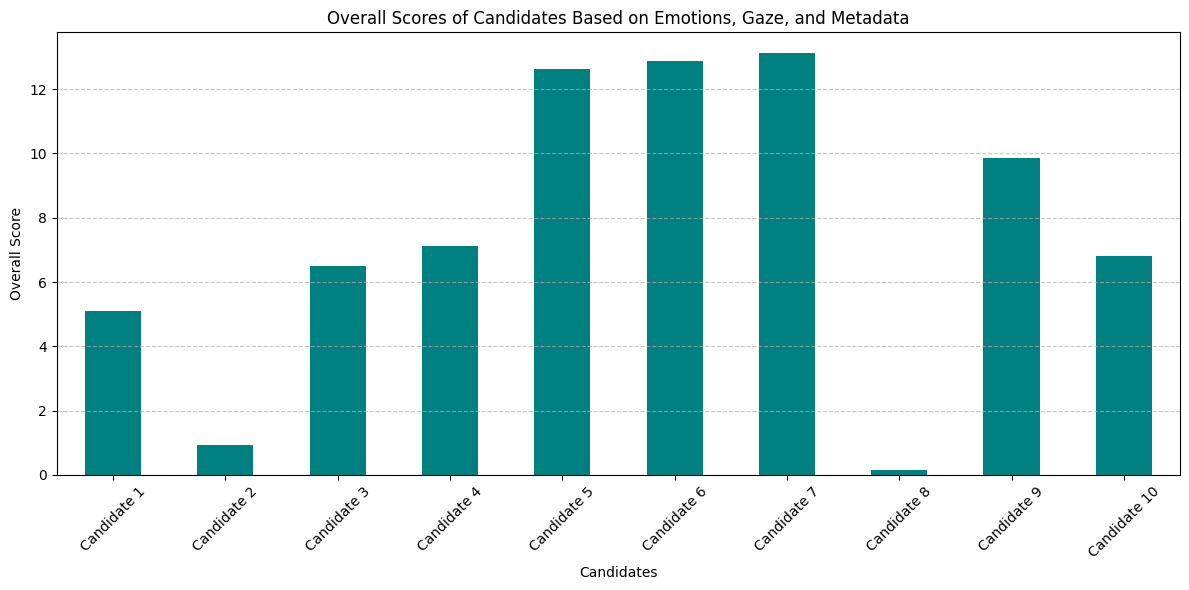
**Analysis of Combined Scores Visualization and Overall Scores**

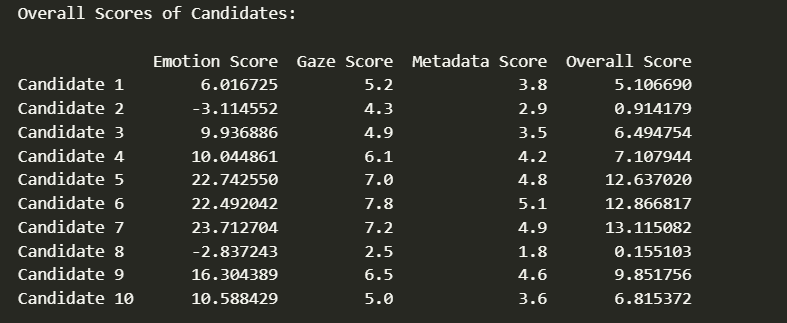
The bar chart presents the overall scores of each candidate based on their combined emotion, gaze, and metadata scores, using weighted averages.



**Observations:**

1. **Top Performers:**
   * **Candidates 5, 6, and 7** have the highest overall scores, with values around 12 to 13. These candidates are strong in all evaluated aspects, making them suitable for recruitment.
   * High scores suggest well-balanced emotions, consistent gaze behavior, and stable metadata performance during their introduction videos.
2. **Moderate Performers:**
   * **Candidates 3, 4, 9, and 10** fall in the mid-range of scores (6 to 9), showing potential but with room for improvement.
   * Their scores suggest balanced performances, but they may benefit from further refining their presentation skills, such as better emotional engagement or improved gaze consistency.
3. **Low Performers:**
   * **Candidates 1, 2, and 8** have the lowest scores, with Candidate 2 and Candidate 8 scoring particularly low.
   * Their lower scores are primarily due to poor emotional expression or inconsistent gaze metrics, which could negatively impact their presentation quality.



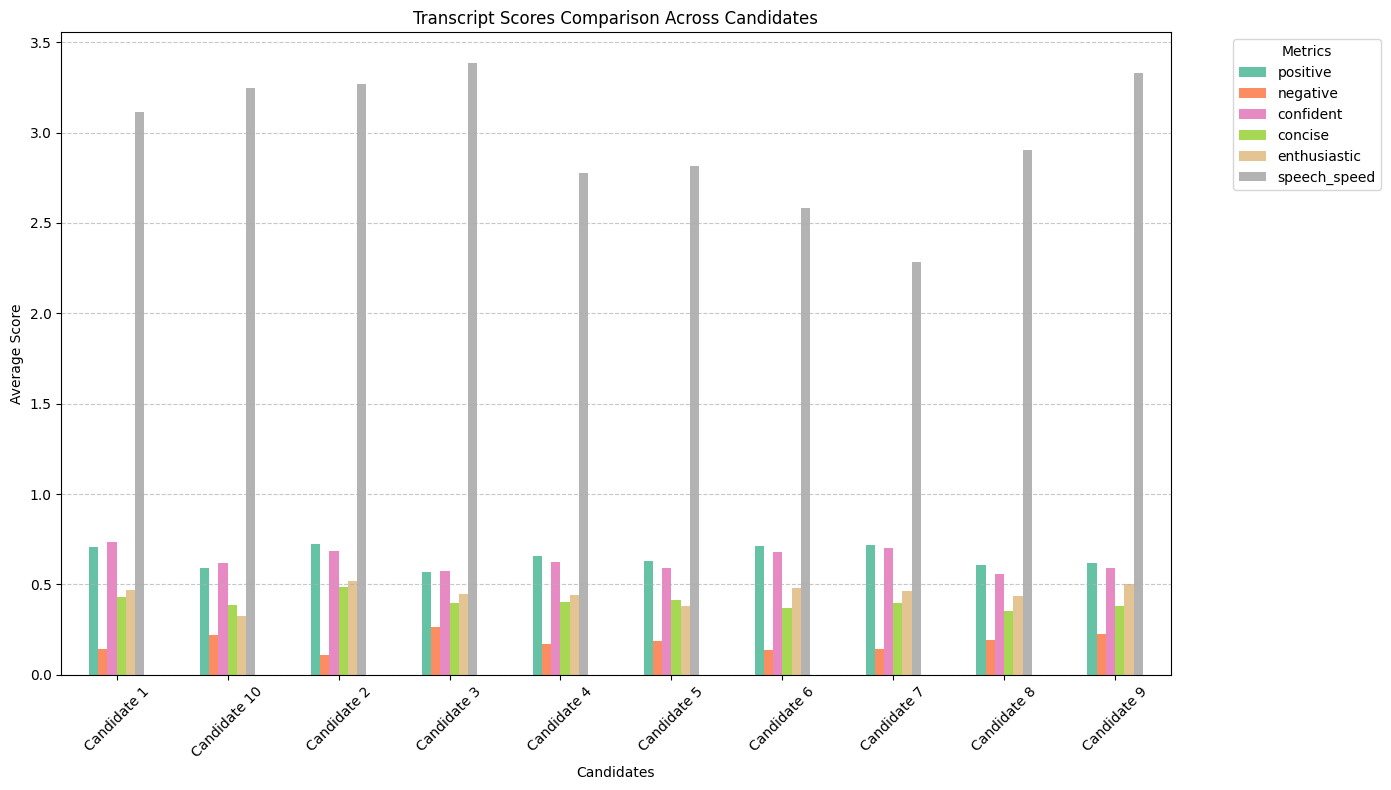


**Analysis of Transcript Scores and Overall Insights**

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

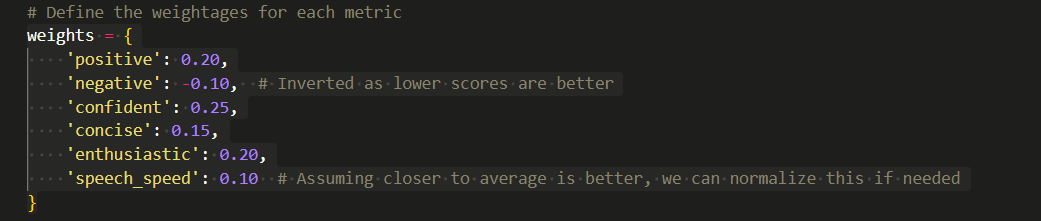
**Transcript Scores Analysis:**

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.



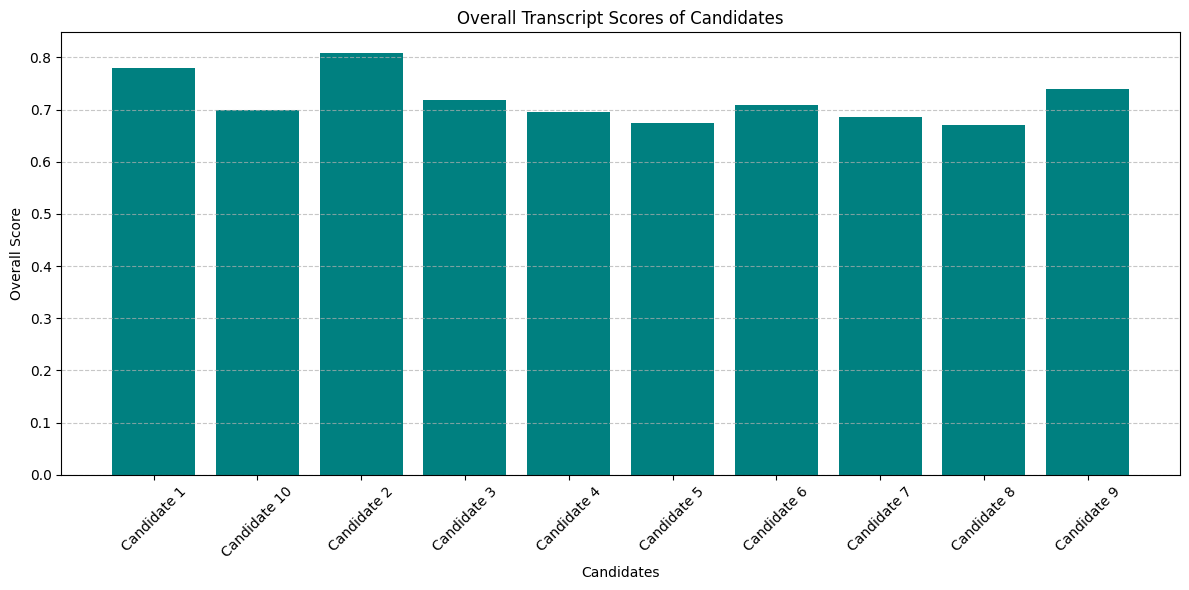
**Analysis of Overall Transcript Scores of Candidates**

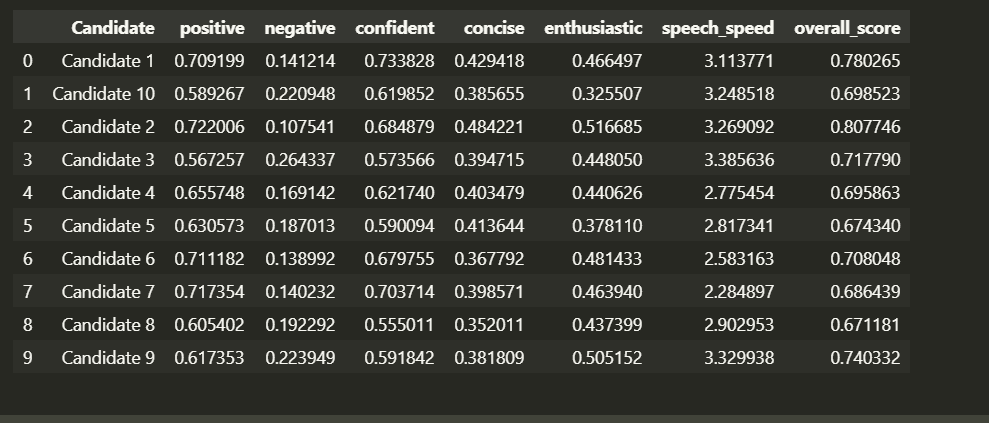
The bar chart illustrates the overall transcript scores of each candidate based on various weighted metrics, including positivity, negativity, confidence, conciseness, enthusiasm, and speech speed.



**Key Observations:**

1. **Top Performers:**
   * **Candidate 2** achieved the highest overall transcript score of 0.807, indicating a well-rounded performance across all evaluated metrics.
   * **Candidate 9** follows closely with a score of 0.740, showing strong positive, confident, and enthusiastic communication skills.
2. **Moderate Performers:**
   * **Candidates 1, 3, and 6** also scored above 0.7, suggesting they performed well in several aspects, particularly confidence and positivity.
   * **Candidate 6**, despite scoring well, has slightly lower conciseness, suggesting opportunities for improvement in making speech more succinct.
3. **Lower Performers:**
   * **Candidate 10** and **Candidate 7** scored around 0.68, which, while decent, indicates room for improvement, particularly in enthusiasm and speech speed control.



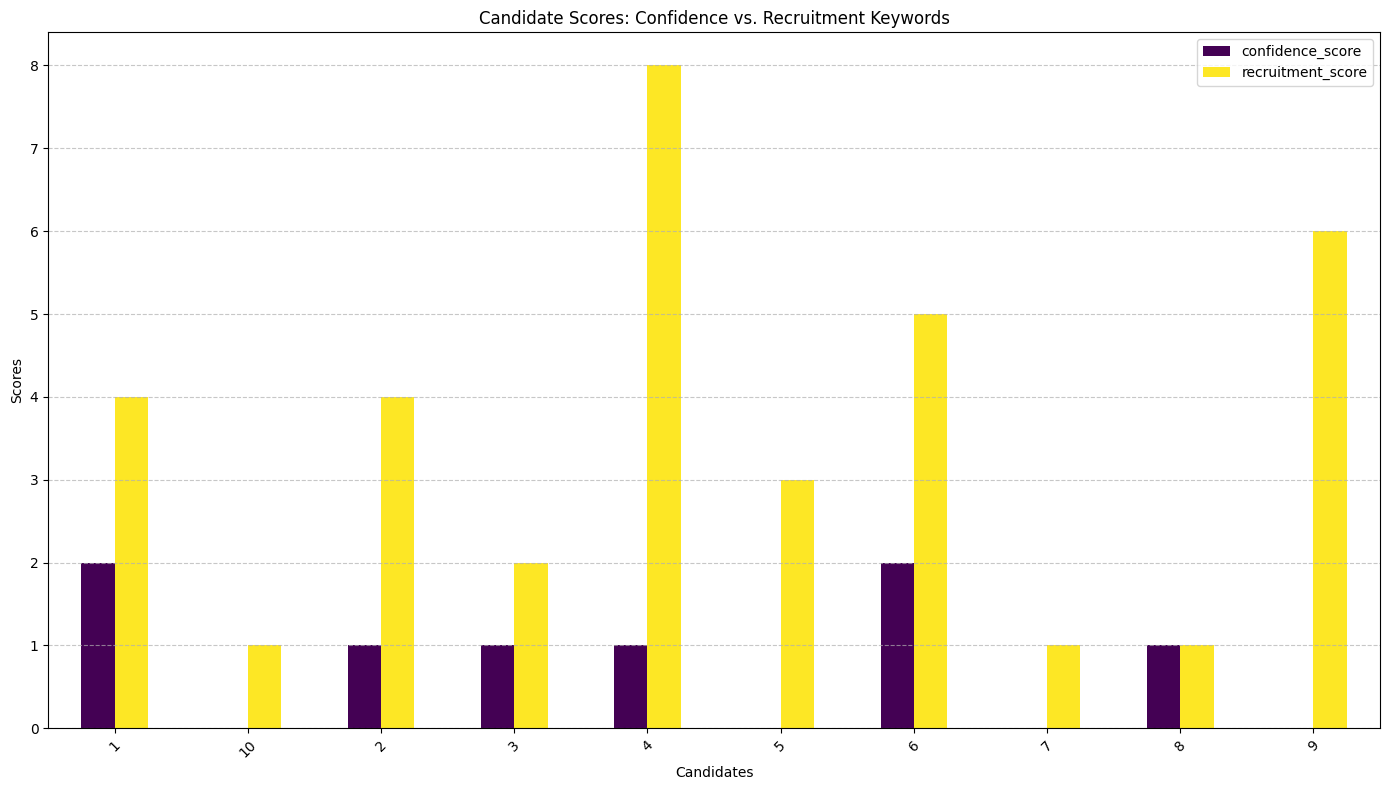


**Analysis of Confidence vs. Recruitment Keywords in Candidate Transcripts**

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.

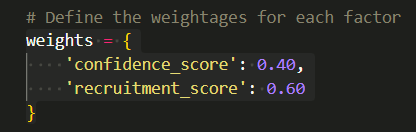
**Key Observations:**

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.



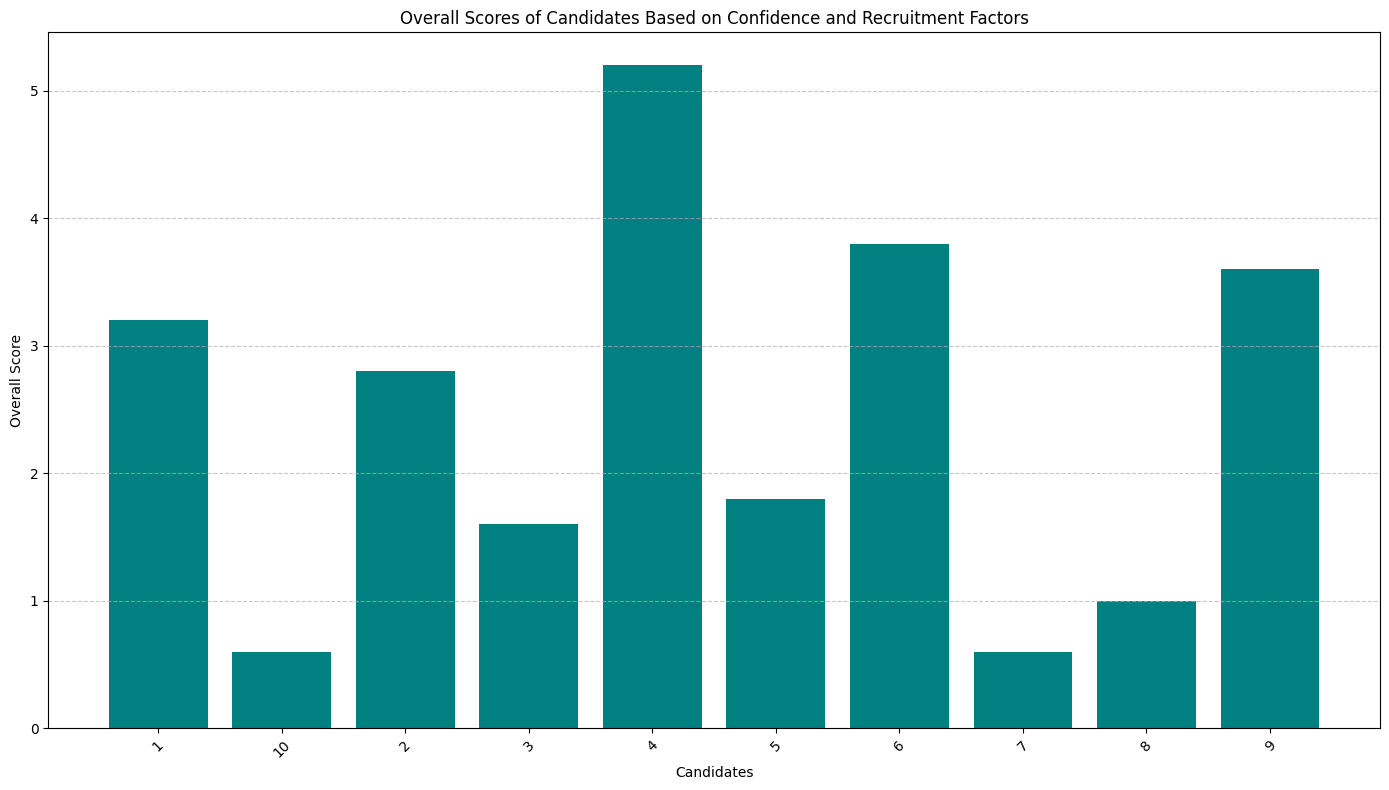
**Analysis of Combined Scores Based on Confidence and Recruitment Factors**

The bar chart shows the overall scores of candidates based on their confidence and recruitment keyword usage, calculated with weights of 0.40 for confidence score and 0.60 for recruitment score.



**Key Observations:**

1. **Top Performers:**
   * **Candidate 4** stands out with the highest overall score of 5.2, driven by strong recruitment keyword usage and moderate confidence. This indicates they effectively communicated their qualifications and skills relevant to recruitment.
   * **Candidates 6 and 9** also scored well, demonstrating a good balance between confidence and the use of key recruitment terms, making them strong contenders.
2. **Moderate Performers:**
   * **Candidates 1 and 2** scored above average due to solid recruitment keyword usage paired with moderate confidence. These candidates show potential but could improve their delivery for a higher impact.
   * **Candidate 5** had a decent recruitment score but lacked confidence, indicating a need for development in how assertively they present their strengths.
3. **Low Performers:**
   * **Candidates 7, 8, and 10** scored the lowest, with minimal use of recruitment keywords and low confidence levels. These scores suggest that these candidates struggled to communicate their qualifications effectively.





**Final Analysis of Combined Scores of Candidates**

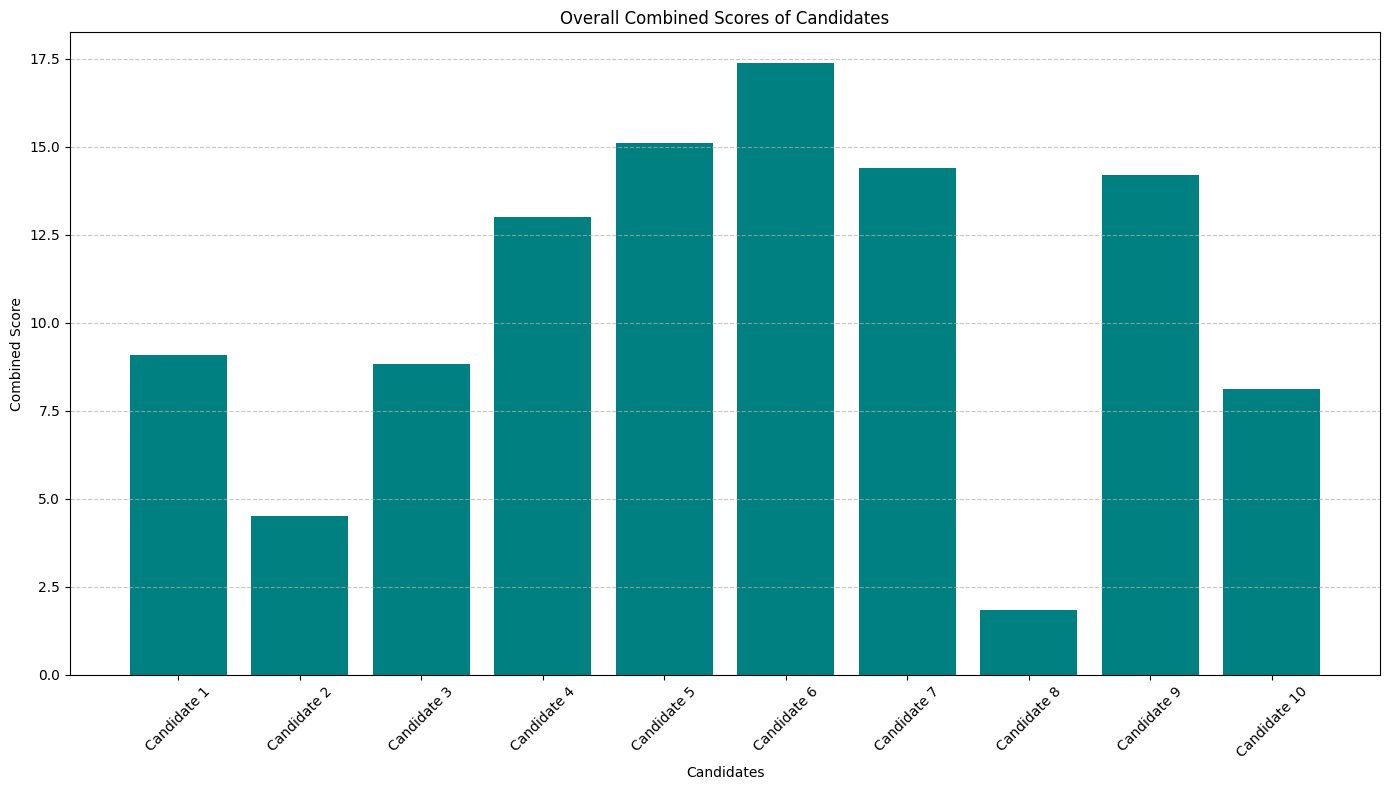
The visualizations and table provide a comprehensive overview of the candidates' overall performance based on their emotion, transcript, and text scores. The combined score reflects each candidate’s overall suitability for recruitment based on these metrics.

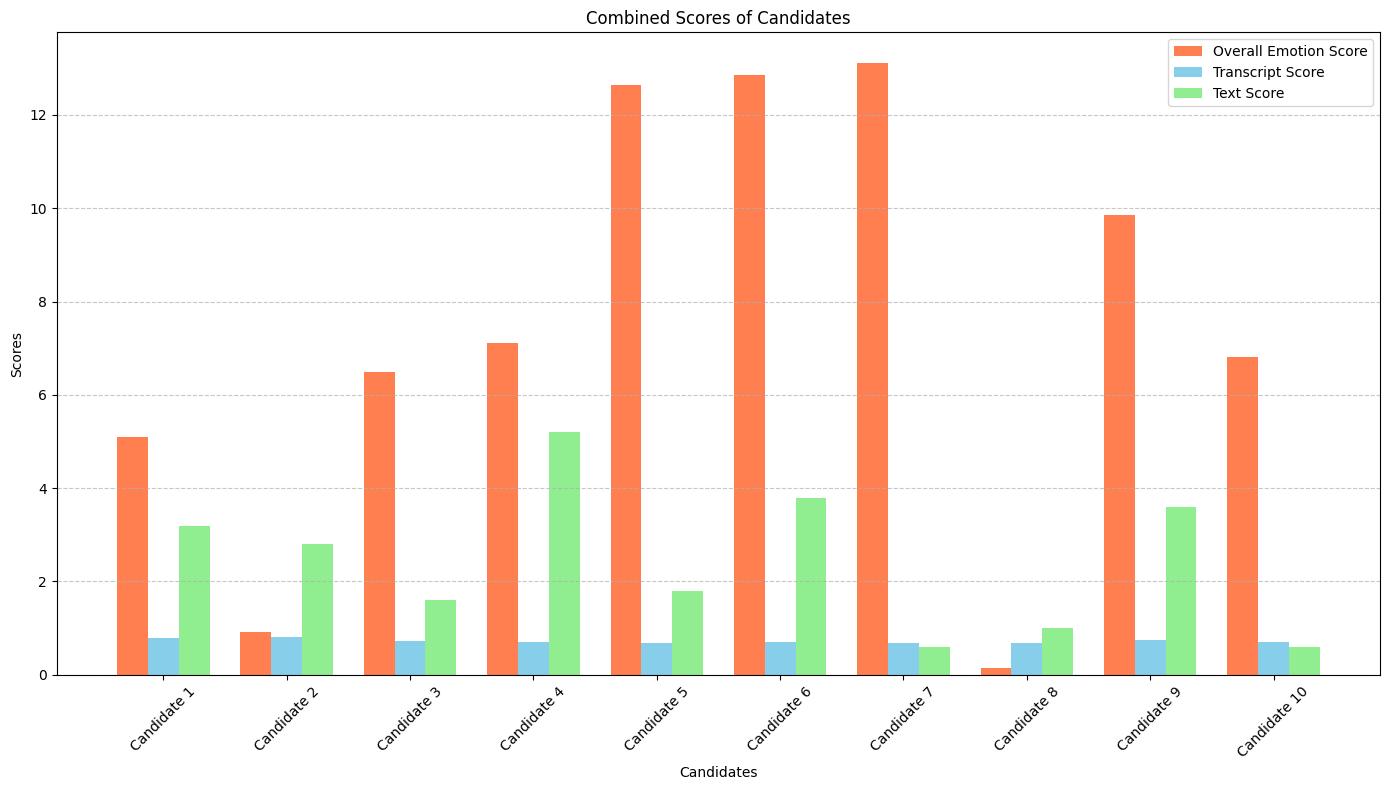
**Key Observations:**

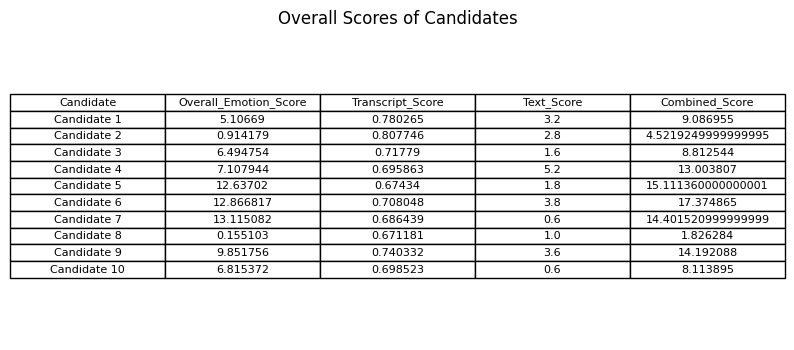
1. **Top Performers:**
   * **Candidate 6** stands out with the highest combined score of 17.37, indicating consistent high performance across all evaluated metrics. This candidate demonstrates a balanced approach in emotional expression, strong transcript delivery, and relevant content in their communication.
   * **Candidates 5, 7, and 9** also perform well with scores above 14, showing strengths in emotional engagement and effective communication skills.
2. **Moderate Performers:**
   * **Candidates 1, 3, and 10** score between 8 and 9, indicating solid, though not exceptional, performances. They show potential but may require specific improvements in either emotional expression or content relevance to enhance their overall impact.
3. **Lower Performers:**
   * **Candidates 2 and 8** have the lowest combined scores, with Candidate 8 scoring particularly low at 1.83. These candidates may lack engagement, confidence, or relevant content in their communication, indicating significant areas for improvement.

**Insights and Recommendations:**

* **Prioritize Recruitment for Top Candidates:** Focus on Candidates 6, 5, 7, and 9, as their high combined scores suggest a well-rounded and impactful presentation style suitable for roles requiring strong communication skills.
* **Development Opportunities for Moderate Performers:** Candidates 1, 3, and 10 would benefit from targeted feedback to refine specific aspects of their performance, such as enhancing emotional engagement or improving the clarity and relevance of their message.
* **Significant Improvement Needed for Low Scorers:** Candidates 2 and 8 need substantial support to boost their communication effectiveness, focusing on building confidence, enhancing emotional expressiveness, and increasing the relevance of their spoken content.







**Actionable Insights Based on Data Analysis**

**1. Recruitment Recommendations:**

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

* **Recommended for Recruitment:**
  + **Candidate 6** (Score: 17.37) - This candidate is highly recommended due to their consistent high scores across all metrics. They show strong emotional engagement, effective communication skills, and relevant content in their presentations.
  + **Candidate 5** (Score: 15.11) - Demonstrates excellent emotional expression and communication clarity. Their presentation style is engaging, making them 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. They are a good fit but may benefit from minor coaching to enhance confidence in communication.
  + **Candidate 1** (Score: 9.09) - Shows good emotional engagement but needs improvement in confidence and refining their 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, demonstrating the ability to engage and connect with their audience effectively.
* **Areas of Expertise:**
  + **Candidate 6:** Strong in technical communication and engaging storytelling, making them ideal for roles that involve explaining complex concepts clearly.
  + **Candidate 5:** Excellent at delivering persuasive messages with an emphasis on leadership and team experience.
  + **Candidate 9:** Effective in conveying strategic and analytical skills, showing expertise in structured presentations.
* **Candidates Needing Improvement:**
  + **Candidate 8:** Struggles with maintaining engagement; needs to develop both content quality and delivery style.
  + **Candidate 2:** Lacks assertiveness; requires coaching to boost confidence and enhance their speaking style.

**3. Additional Insights:**

* **Engagement Strategies:**
  + Candidates with high scores showed a pattern of using storytelling and personal anecdotes, which could be integrated into training for lower-performing candidates to enhance engagement.
* **Confidence Building:**
  + For candidates with lower confidence scores, consider offering workshops focusing on public speaking, body language, and vocal training to improve presentation skills.
* **Content Relevance:**
  + Emphasize the importance of using relevant keywords and structuring content logically to make presentations more impactful. Tailored feedback sessions can help candidates integrate this into their speech delivery.