PROMPT ENGINEERING

DOCUMENTATION



**Name: Vasupally Tarun Dorai Raj**

**Email: Vasupally.dorairaj.21063@iitgoa.ac.in**

Contents

[BASIC STEPS FOLLOWED: 3](#_Toc146830156)

[PROCESS FLOW FOR TRANSCRIPT\_DATA 3](#_Toc146830157)

[PROCESS FLOW FOR TRANSCRIPT TEXT: 10](#_Toc146830158)

[PROMPT AND OUTPUT FOR GAZE.CSV FILES: 17](#_Toc146830159)

[PROMPT AND OUTPUT FOR EMOTION.CSV FILES: 25](#_Toc146830160)

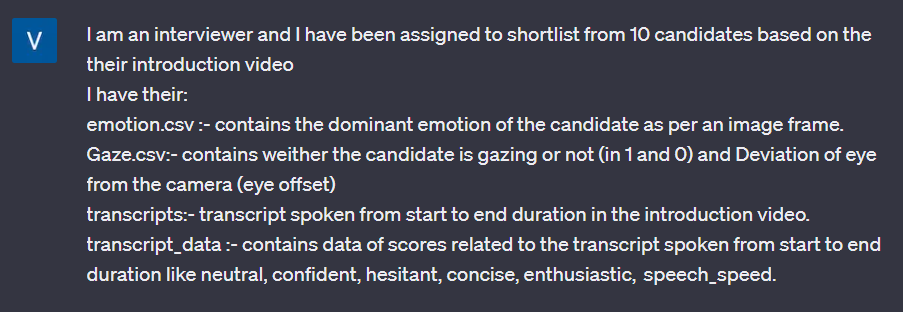
# BASIC STEPS FOLLOWED:

* Explain the situation
* Stick to standardization as the AI model is trained more on the standardized grammar
* Clear instructions should be given
* Adopting a persona
* Specifying the format
* Avoid leading the answer
* Limit the scope

# PROCESS FLOW FOR TRANSCRIPT\_DATA

(text color orange means it is generated by ChatGPT)

* Firstly:-



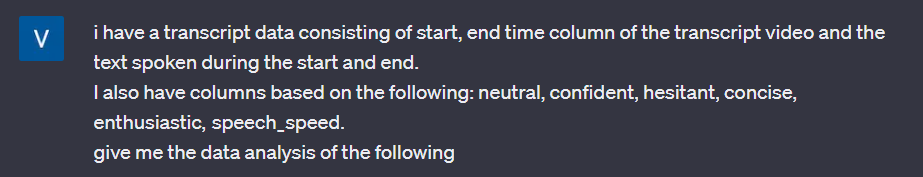
generated:-

It sounds like you have a variety of data related to the candidates' introduction videos that you can use to shortlist them. To make an informed decision, you can follow these steps:

1. **Data Cleaning and Preprocessing**:
   * Ensure that all the data is properly formatted and consistent.
   * Handle missing values if any.
2. **Emotion Analysis (emotion.csv)**:
   * Calculate the dominant emotion for each candidate based on the image frames.
   * Consider whether the dominant emotion aligns with the qualities you're looking for in a candidate. For example, confidence and enthusiasm may be positive emotions.
3. **Gaze and Eye Offset (Gaze.csv)**:
   * Examine whether candidates maintain eye contact (gazing).
   * Analyze the deviation of the eye from the camera (eye offset). Candidates with minimal eye offset may appear more engaged.
4. **Transcript Analysis (transcripts and transcript\_data)**:
   * Analyze the spoken content of each candidate's introduction using the transcripts.
   * Use the transcript\_data to evaluate specific aspects like neutrality, confidence, hesitancy, conciseness, enthusiasm, and speech speed.
   * Assign scores or weights to these aspects based on your evaluation criteria. For example, confident and enthusiastic speech may be more desirable.
5. **Aggregate Scores**:
   * Combine the scores from the emotion analysis, gaze analysis, and transcript analysis to create an overall score for each candidate.
   * You can assign different weights to each aspect based on their importance to the role you're hiring for.
6. **Rank Candidates**:
   * Rank the candidates based on their overall scores. Higher scores indicate stronger candidates.
   * You can use a simple weighted sum or a more complex scoring model, depending on your preferences.
7. **Set a Threshold**:
   * Decide on a threshold score above which candidates will be shortlisted.
8. **Shortlist Candidates**:
   * Shortlist the candidates whose overall scores meet or exceed the threshold.
9. **Review and Finalize**:
   * Review the shortlisted candidates and consider any additional factors that may not be captured in the data (e.g., experience, qualifications).
   * Finalize the list of candidates to be invited for further interviews or assessments.

Remember that the specific weights and criteria you use for scoring should align with the requirements of the position you're hiring for. Additionally, it's essential to remain objective and consistent in your evaluation to ensure a fair selection process.

* Here we express our persona to ChatGPT to understand the situation and generate results according to that. We also explain the data given to us so that it gives us ideas that could be followed and are useful for the next prompts.

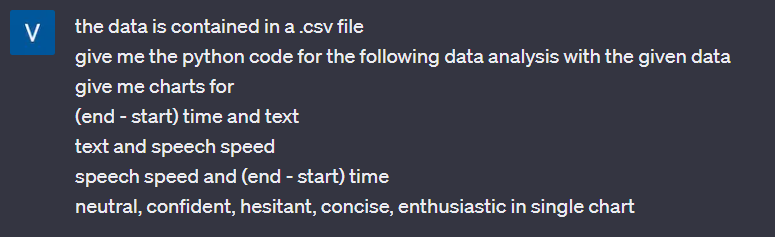


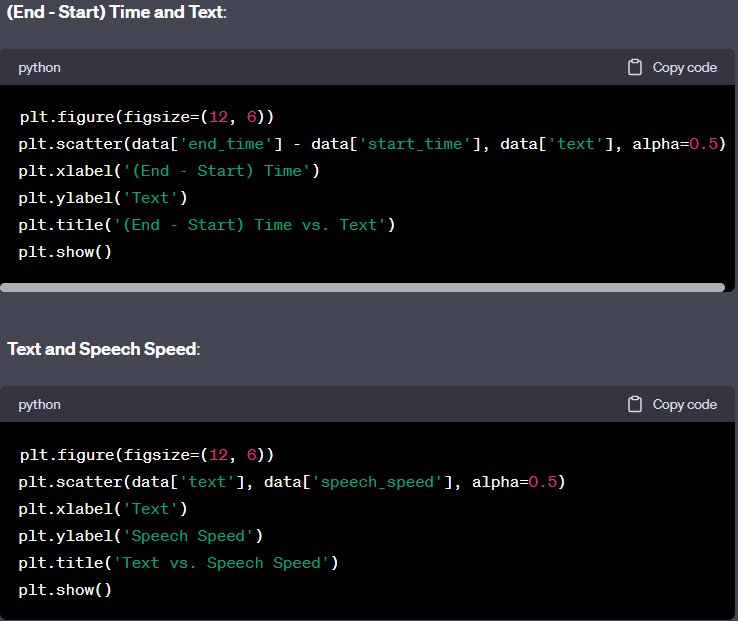
Analyzing transcript data with columns for start and end times, text spoken, and attributes like neutral, confident, hesitant, concise, enthusiastic, and speech speed can provide valuable insights into the spoken content. Here are some data analysis tasks you can perform on this data:

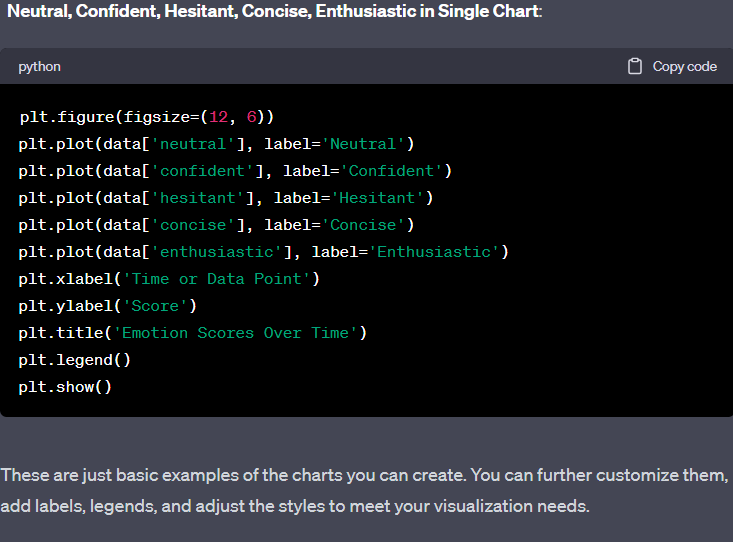
1. **Descriptive Statistics**: Compute basic statistics for each attribute, such as mean, median, minimum, maximum, and standard deviation. This will give you an overview of the data's distribution.
2. **Sentiment Analysis**: Analyze the text spoken during each interval to determine the predominant sentiment (neutral, confident, hesitant, enthusiastic, etc
3. **Speech Speed Analysis**: Calculate the average speech speed (words spoken per minute) for each interval. This can help identify fast-talking or slow-talking segments.
4. **Duration Analysis**: Calculate the duration of each interval (end time - start time) and analyze the distribution of interval durations. This can help identify long or short segments.
5. **Correlations**: Investigate potential correlations between attributes. For example, you might want to see if there is a correlation between confidence level and speech speed.
6. **Trends Over Time**: Plot how attributes change over time. For example, visualize how speech speed varies throughout the video.
7. **Keyword Analysis**: Identify frequently occurring keywords or phrases in the spoken text. This can provide insights into the main topics or subjects discussed.
8. **Visualizations**: Create visualizations, such as line plots, bar charts, or heatmaps, to present the analysis results in a clear and interpretable manner.

Remember that the specific analysis tasks will depend on your research objectives and the nature of the transcript data. You may need to use a combination of data analysis tools, NLP libraries, and visualization libraries (e.g., pandas, NLTK, spaCy, Matplotlib, seaborn) to perform these tasks effectively.

* Now I have specified the output format and explained the data given in the above prompt.

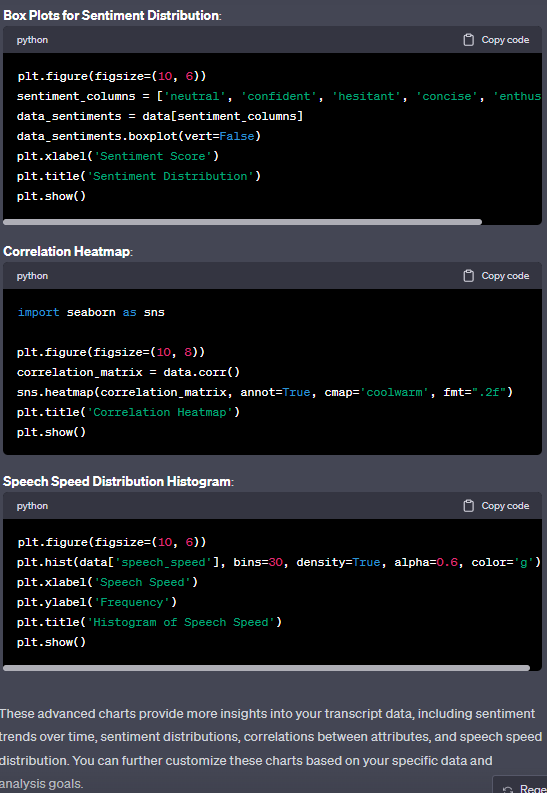




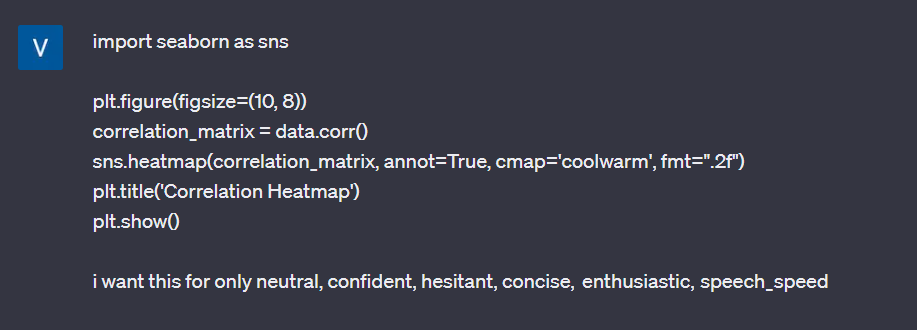


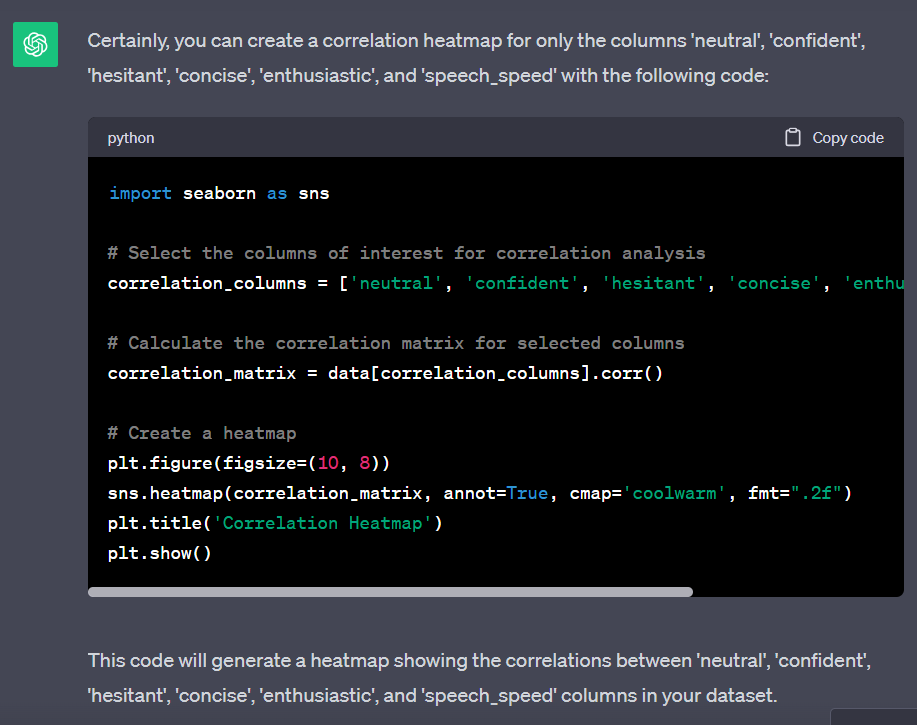
* For further plotting of similar graphs in a different way:





* I wanted the correlation heatmap for only emotions like confident, neutral, hesitant, concise, enthusiastic, speech speed.

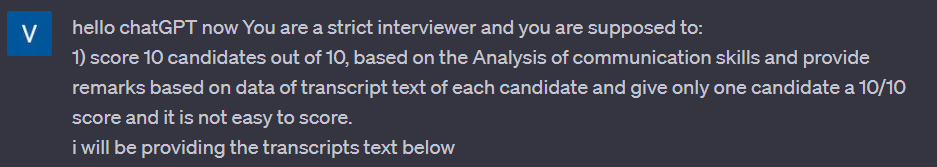




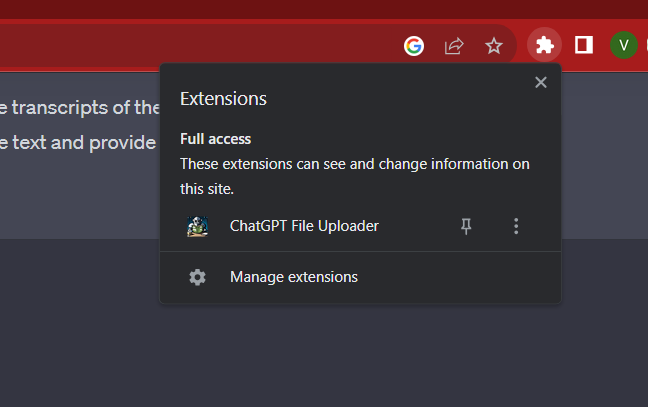
# PROCESS FLOW FOR TRANSCRIPT TEXT:

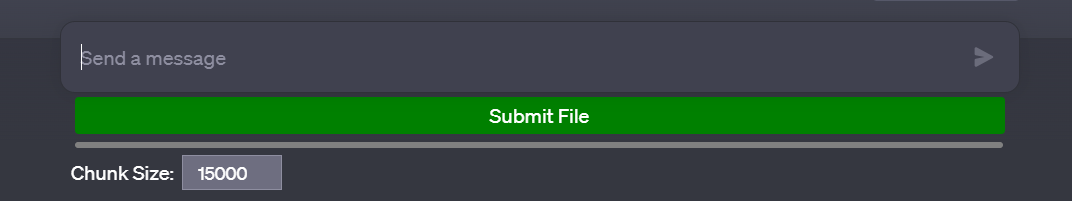
(Analysis of communication skills and finding areas of expertise)

* Prompt:

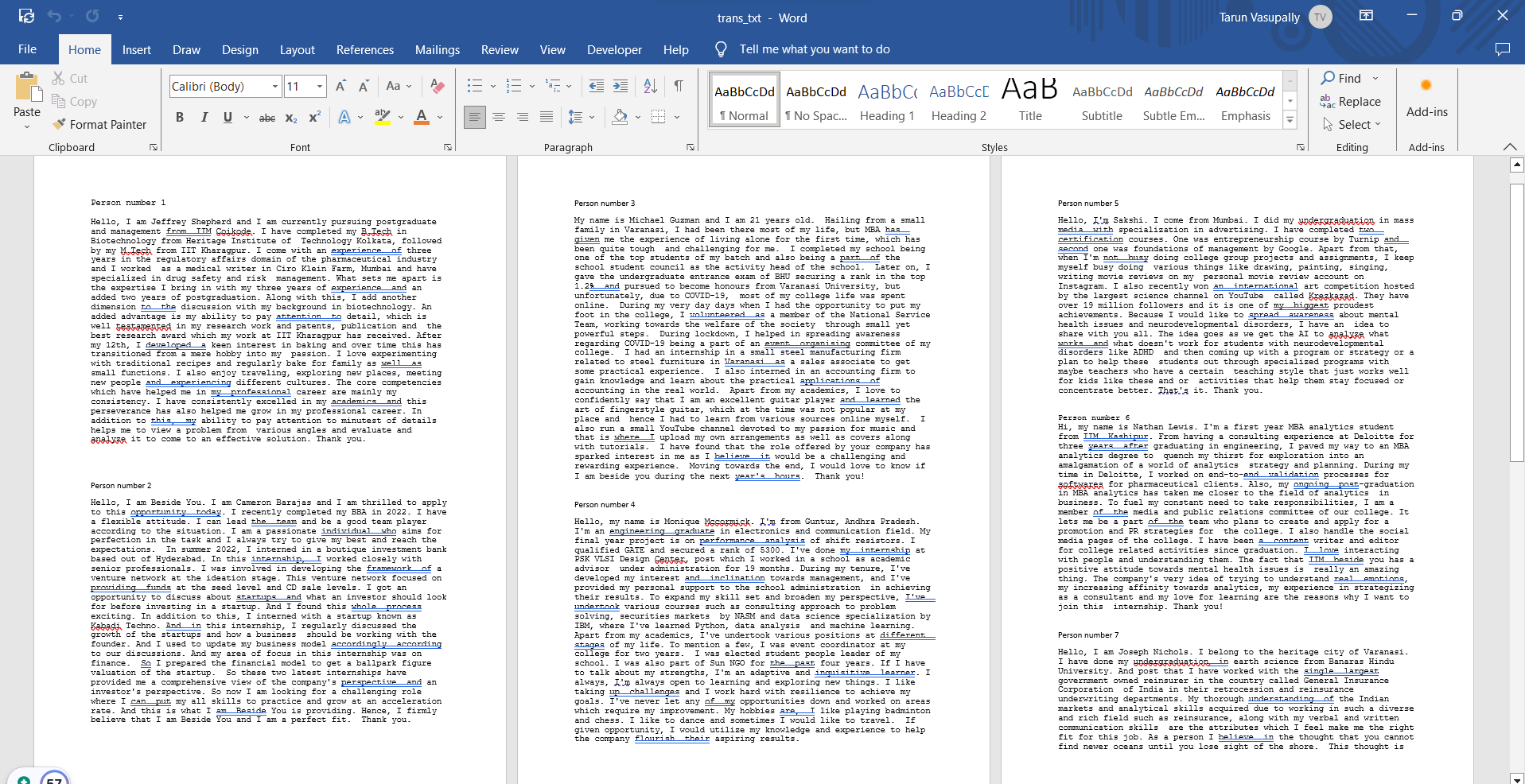


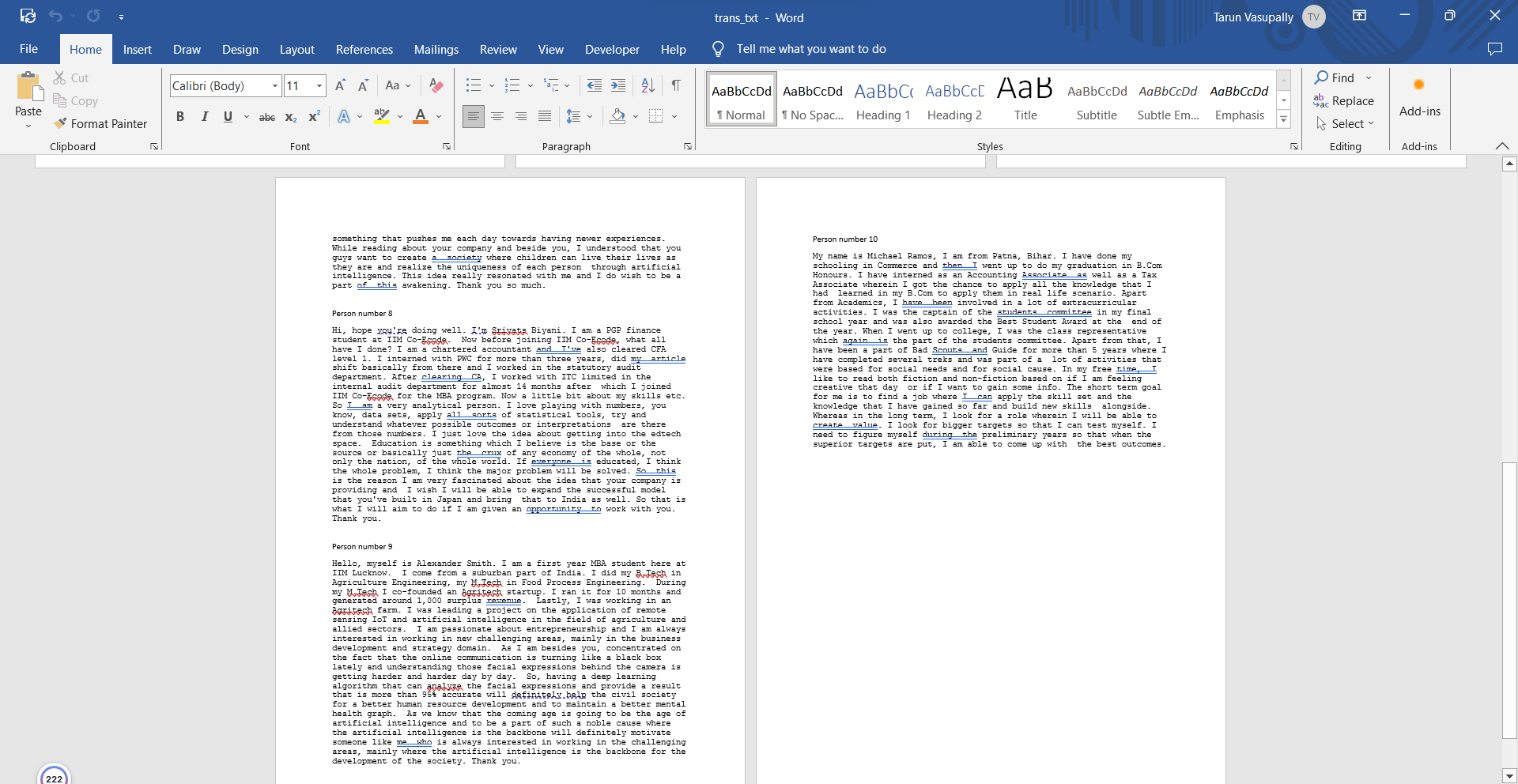
Extension (for submitting large files that takes in the data in the file)





Here I have submitted the docx file that contains all the candidate’s transcript text. Below are the document images:





Generated output via prompt given:

Candidate 1:

Jeffrey Shepherd:

Score: 9/10

Remarks: Jeffrey Shepherd demonstrates strong communication skills and provides a detailed overview of his education, work experience, and personal interests. He emphasizes his ability to set and achieve goals, which adds to his profile positively.

Candidate 2:

Cameron Barajas:

Score: 7/10

Remarks: Cameron Barajas presents a concise overview of his qualifications and internship experiences. While he mentions his flexibility and passion, his response could benefit from more specific examples of setting and achieving targets.

Candidate 3:

Michael Guzman:

Score: 6/10

Remarks: Michael Guzman shares his academic and extracurricular background but could improve by providing concrete examples of goal-setting and achievement. His response still lacks a strong connection to the job.

Candidate 4:

Monique McCormick:

Score: 8/10

Remarks: Monique effectively highlights her educational background and work experiences. While she mentions her passion for challenges and resilience, more specific examples of setting and exceeding targets would enhance her response.

Candidate 5:

Sakshi:

Score: 8/10

Remarks: Sakshi provides a good overview of her academic background and certifications. Her mention of winning an international art competition demonstrates a commitment to achieving goals. She could elaborate more on her goals related to using AI for neurodevelopmental disorders.

Candidate 6:

Nathan Lewis:

Score: 8/10

Remarks: Nathan effectively communicates his transition from consulting to pursuing an MBA in analytics. He mentions his desire for exploration and highlights his involvement in college committees. His enthusiasm for setting and achieving targets is evident.

Candidate 7:

Joseph Nichols:

Score: 7/10

Remarks: Joseph shares his educational and work background and expresses a desire for new experiences. While he mentions setting targets, his response could benefit from more specific examples of goal achievement.

Candidate 8:

Srivats Biyani:

Score: 8/10

Remarks: Srivats provides a clear overview of his educational and professional journey. He expresses a strong interest in education and setting bigger targets. He could elaborate further on his approach to achieving these targets.

Candidate 9:

Alexander Smith:

Score: 9/10

Remarks: Alexander effectively communicates his background and passion for entrepreneurship and AI. He links his experiences to the role's focus on AI and facial expression analysis. His commitment to setting and achieving bigger targets aligns with the company's mission.

Candidate 10:

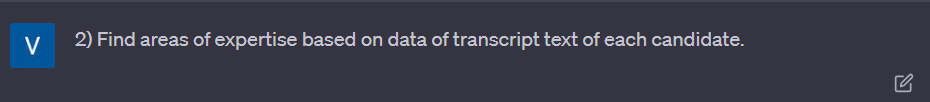
Michael Ramos:

Score: 8/10

Remarks: Michael Ramos presents a well-rounded profile, highlighting his educational background in Commerce and his internship experience in accounting and taxation.

I would maintain the score of **10/10 for Alexander Smith** for his clear articulation of qualifications, relevant experience, and passionate alignment with the company's mission, including setting and achieving bigger targets.

* Next, we needed to find the area of expertise for each of the candidates so we entered the following prompt and here it takes the data that has been submitted which is the docs file.



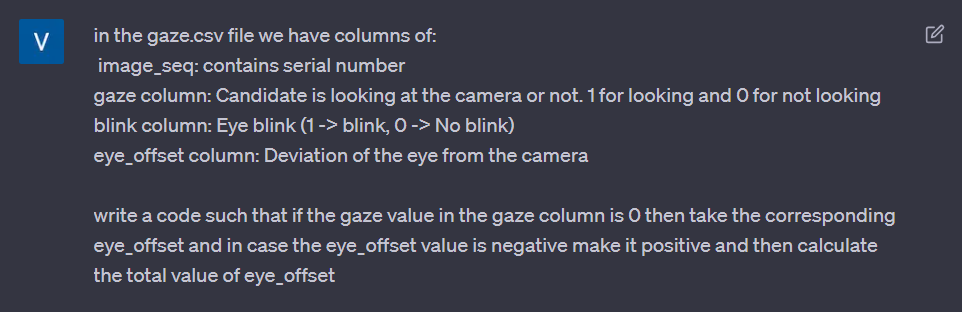
Based on the transcript text provided for each candidate, here are the areas of expertise that can be inferred:

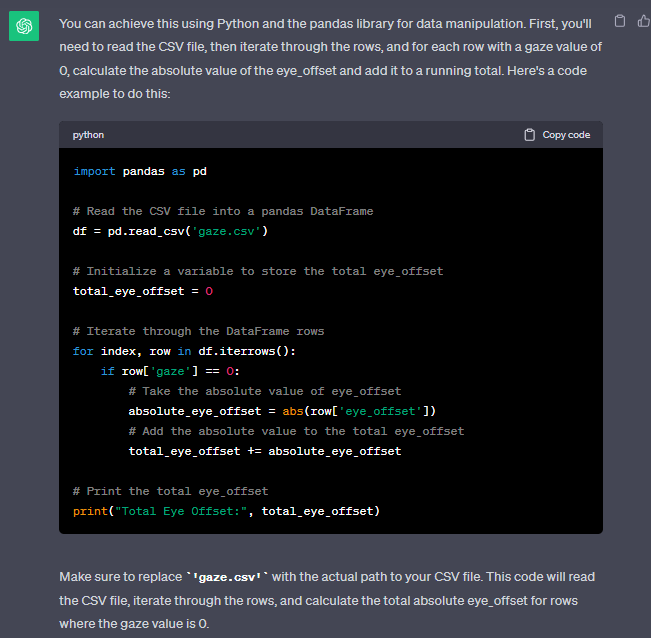
1. **Jeffrey Shepherd**:
   * Areas of Expertise: Regulatory affairs in the pharmaceutical industry, medical writing, drug safety, risk management, biotechnology, baking, research, attention to detail.
2. **Cameron Barajas**:
   * Areas of Expertise: Venture network development, startup growth, financial modeling, finance.
3. **Michael Guzman**:
   * Areas of Expertise: Academic excellence, event organizing, volunteering, sales, accounting, guitar playing, fingerstyle guitar, online content creation.
4. **Monique McCormick**:
   * Areas of Expertise: Electronics and communication engineering, academic advising, auditing, data analysis, machine learning, badminton, chess.
5. **Sakshi**:
   * Areas of Expertise: Mass media, advertising, entrepreneurship, art (drawing and painting), singing, writing, mental health awareness.
6. **Nathan Lewis**:
   * Areas of Expertise: Consulting, validation processes, softwares for pharmaceutical clients, media and public relations, social media management, content writing.
7. **Joseph Nichols**:
   * Areas of Expertise: Earth science, reinsurance, analytical skills, market understanding, adaptability to new experiences.
8. **Srivats Biyani**:
   * Areas of Expertise: Chartered accountancy, CFA, internal audit, financial analysis, edtech, finance, data analysis.
9. **Alexander Smith**:
   * Areas of Expertise: Agriculture engineering, entrepreneurship, agritech startups, remote sensing, IoT, artificial intelligence, business development, strategy.
10. **Michael Ramos**:
    * Areas of Expertise: Commerce, accounting, taxation, leadership (students committee, class representative), extracurricular activities (Bad Scouts and Guide), reading, goal setting.

Please note that these areas of expertise are inferred from the provided transcript text.

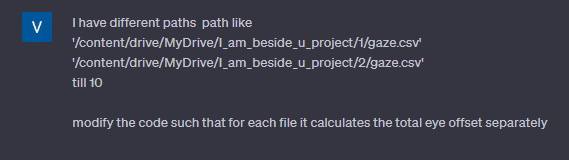
# PROMPT AND OUTPUT FOR GAZE.CSV FILES:

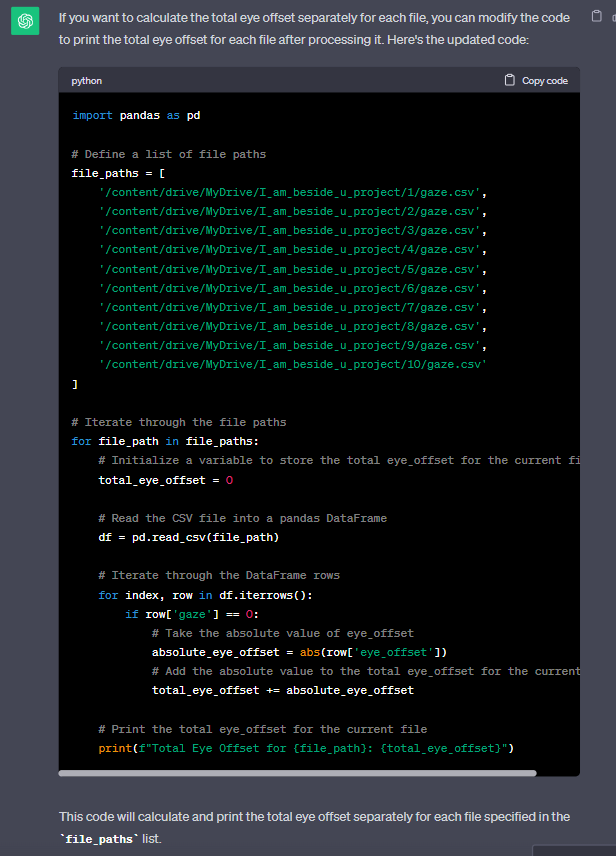
* Here, we have explained to ChatGPT that the gaze.csv file contains many columns like image\_seq, blink, eye\_offset and explained what exactly it means for the model to understand and then specified the format of the output.
* We wanted to know the total eye offset of candidates by taking the corresponding eye\_offset value when the gaze value is 0 and then summing it up. A candidate with the least eye offset total is more likely to be selected.



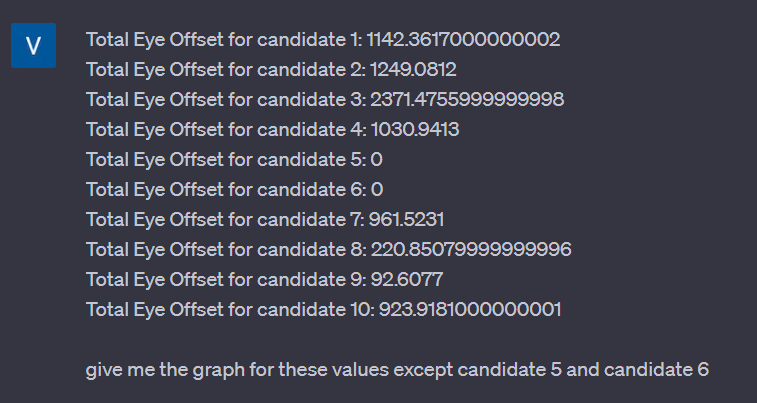


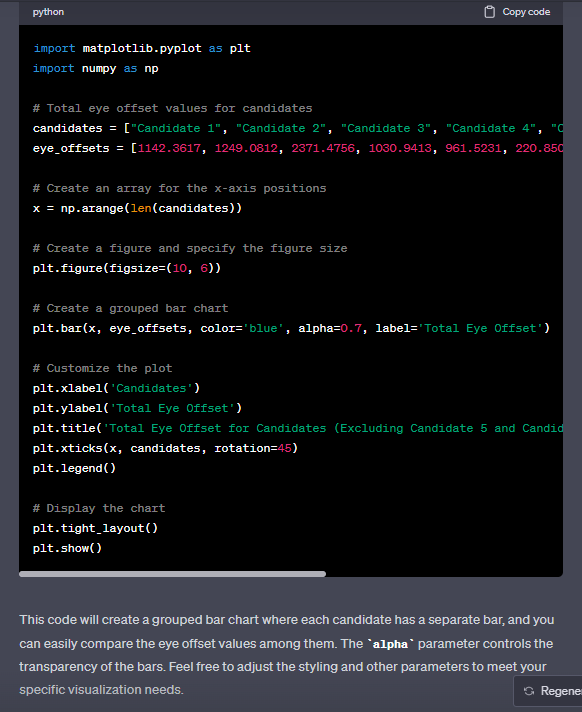
* I had 10 paths for each of the file save, so as to run it in a single code, I used the following prompt

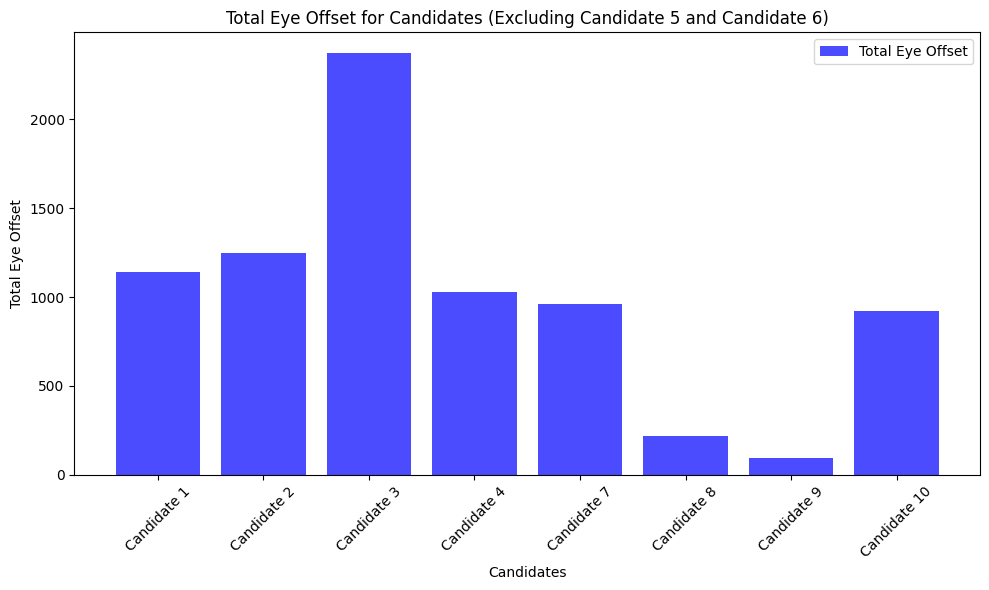




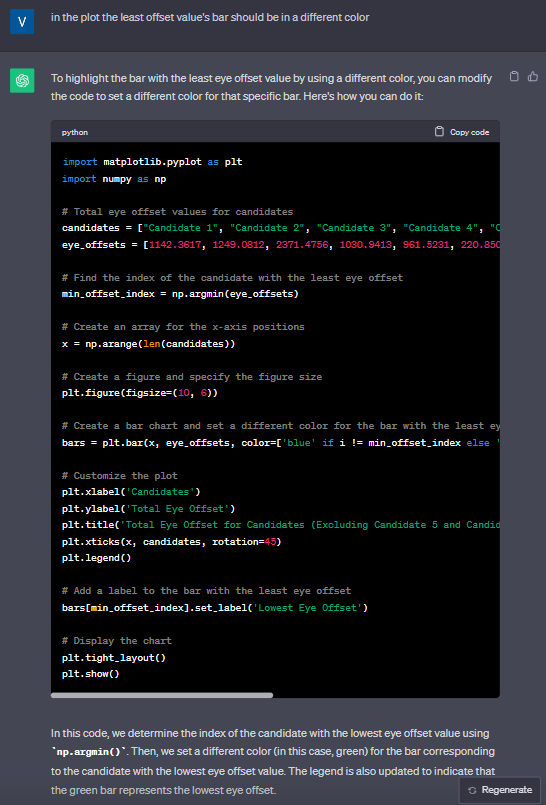
* we have taken the output that we got from the compiler and asked it to plot a graph for easy analysis of the data.
* Notice that, due to insufficient data candidates 5 and 6 got a total eye offset to be 0.
* Thus, we won't be taking this analysis for our candidate selection but if this attribute is a must for screening then assign the mean value of the data to candidates 5 and 6 and then continue with the selection process.

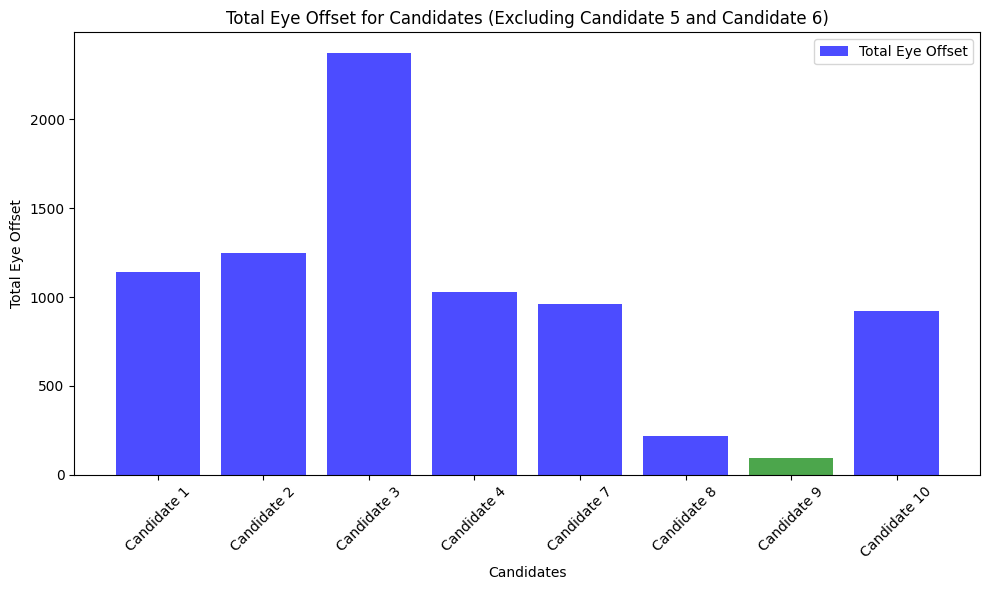






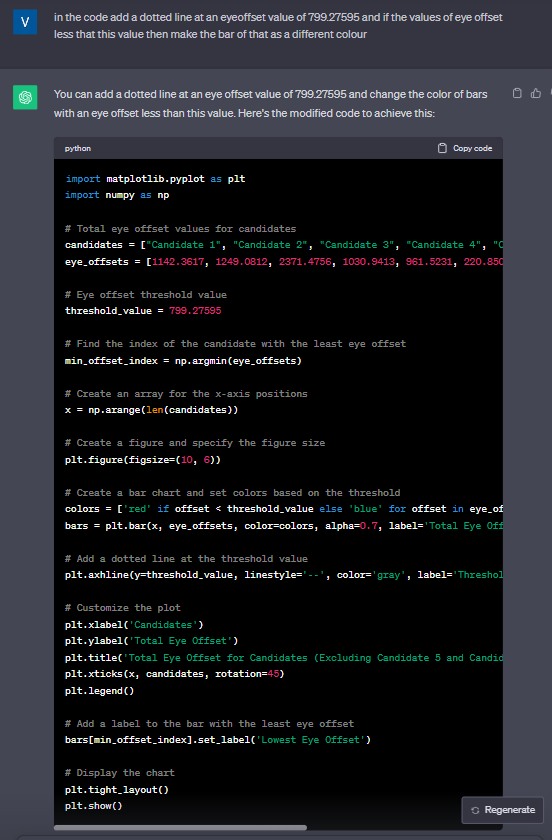
* To understand this data analysis through plots we have entered a prompt such that the candidate with the least offset value has a bar of a different color to easily identify the candidate with the least offset value.





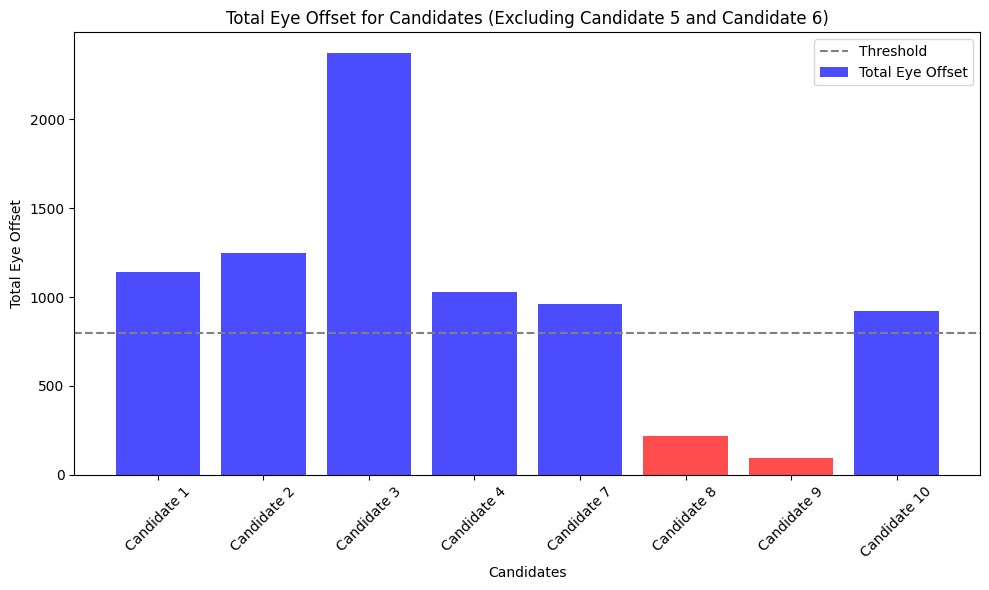
Here we can observe that candidate 9 has the least offset value.

* For further analysis through graphs we can add a threshold value (here we have taken it to be the mean value) and further shortlist the candidates.



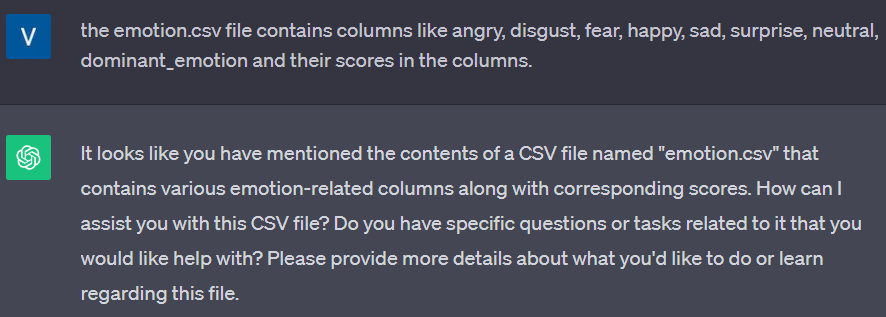
Here 799.27595 is the average value of all the candidate’s eye offset values (candidates 5 and 6 are not considered due to insufficient data)

If candidate 5 and 6 are to be considered then assign them the average values.

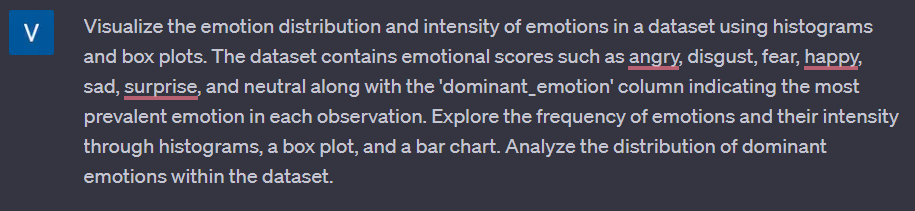


# PROMPT AND OUTPUT FOR EMOTION.CSV FILES:

* Firstly, we provide information on the emotion.csv file that it contains columns of scores of emotions like angry, disgust, fear, happy, sad, surprise, neutral, and dominant\_emotion.



* Here we have given different emotions for which we want the histogram plot code, boxplot, and dominant emotion for each of the following emotion.csv data.





We have modified the code like adding file paths and adjusting the bins sizes.