EXPLORATORY DATA ANALYSIS REPORT

- Ansh Vardhan

INTRODUCTION:

The report explains the analysis of the data given in the question statement more precisely the gaze.csv , emotion.csv and 1.csv (eg. for person 1) data . I will continue explaining the EDA with the example of Person 1 and the rest of the people have the same plots and charts except the data is altered according to the person which is explained in detail in their respective Jupyter Notebook.

DATASETS USED:

Throughout the EDA I have worked on three different datasets , namely:

1. gaze.csv
2. emotion.csv
3. speech.csv(1.csv)

The EDA has been divided into three parts which individually analyze each of the datasets mentioned above by using different plots and charts and making inferences on them.

Throughout the analysis following libraries have been used:

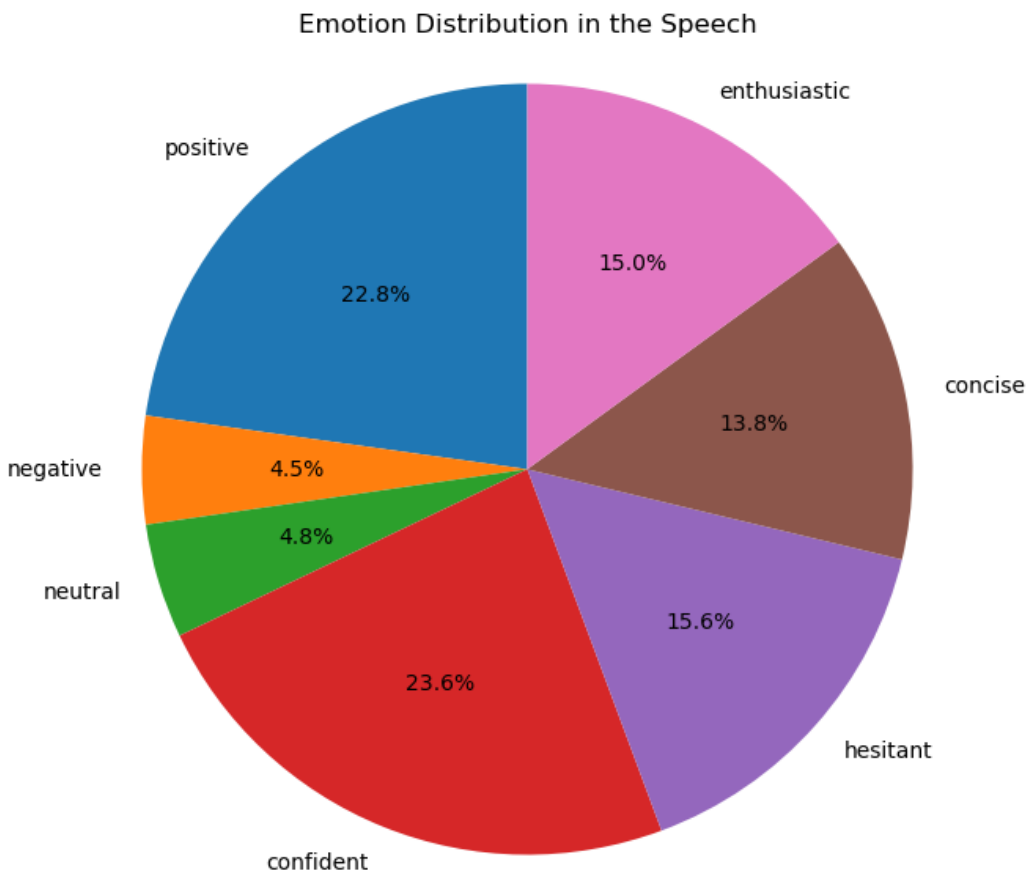
1. Pandas
2. matplotlib.pyplot
3. seaborn

ANALYZING THE SPEECH DATASET:

DESCRIPTIVE ANALYSIS:

In this section we use a Pie chart which describes the distribution of various emotions during the speech.

The pie chart helps in determining the dominant emotions in the speech of the candidate.

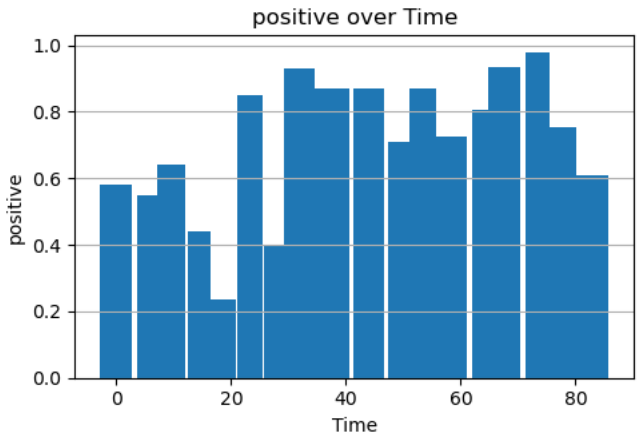


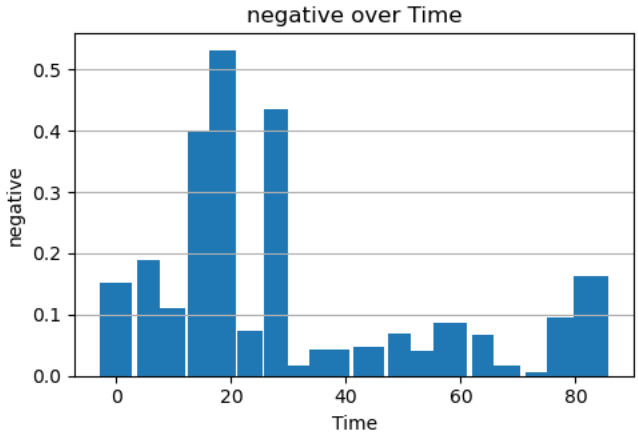
In this case the dominant emotions are:

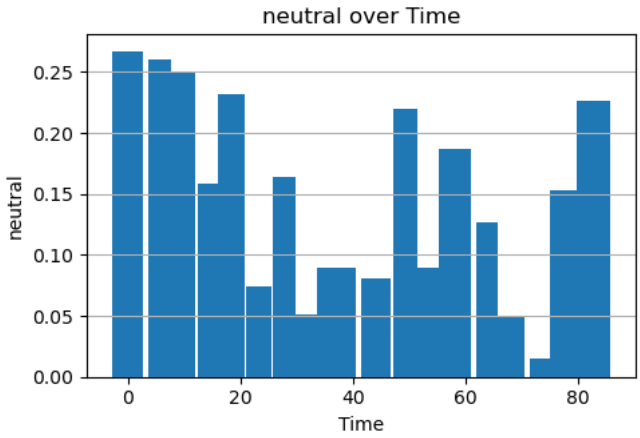
1. Confident
2. Positive
3. Hesitant

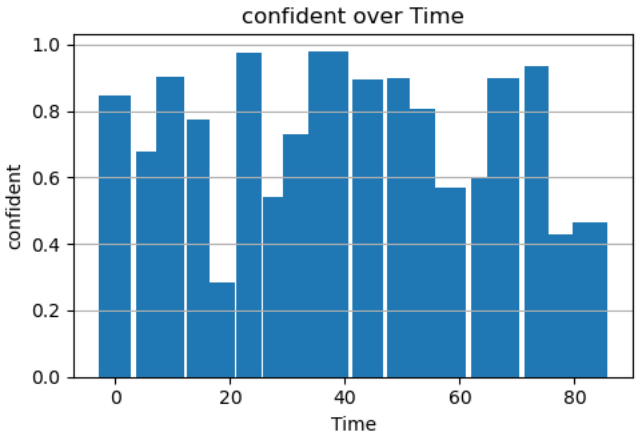
TREND ANALYSIS:

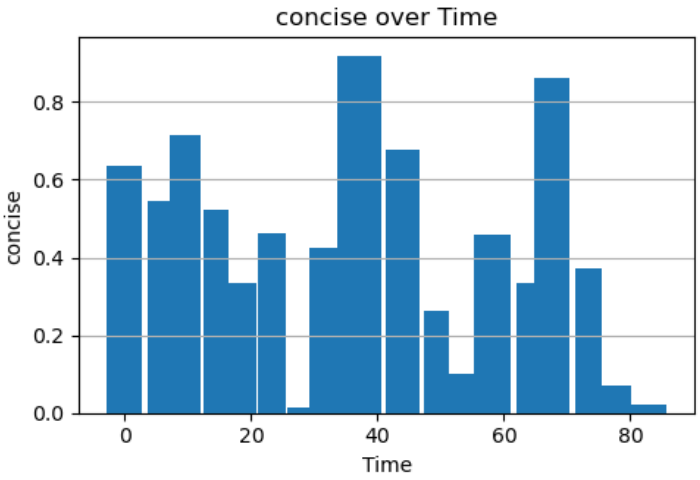
In this section we see the variation of different emotions assessed in speech over time. I have plotted all the emotions in different graphs so that we can get a better understanding of various emotions throughout the speech.

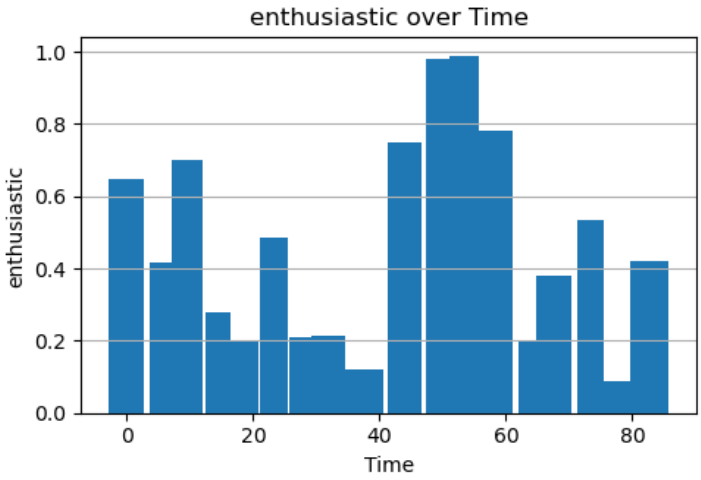


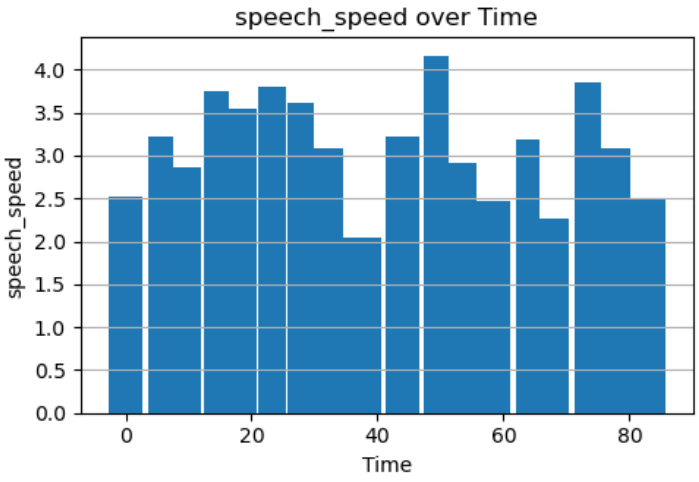


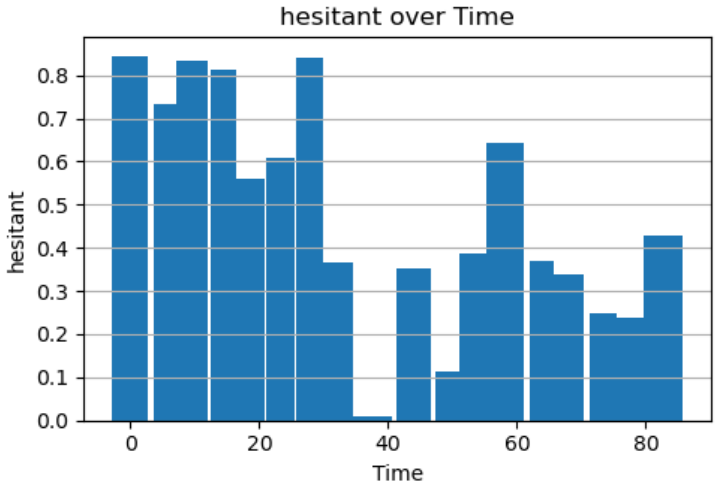












From the graphs, it's evident that the candidate's emotions fluctuate throughout the speech. Using the transcript text and the above graphs we can pinpoint which parts of the speech trigger specific emotional changes by looking at the time where some peak or valley occurs.

By the above process we made the following inferences:

1. The tone of Person1 as a whole sounds more on the positive side than on the negative side as it is evident by the constant higher bars in Positive graph when compared to Negative graph.

2. The Person sounds confident throughout the speech except where he talks about how he worked as a medical writer in Ciro Klein Farm, Mumbai.

3. There is also a dip in confidence when he talks about his ability to pay attention to minutest of details and to view a problem from various angles and evaluate and analyze it to come to an effective solution.

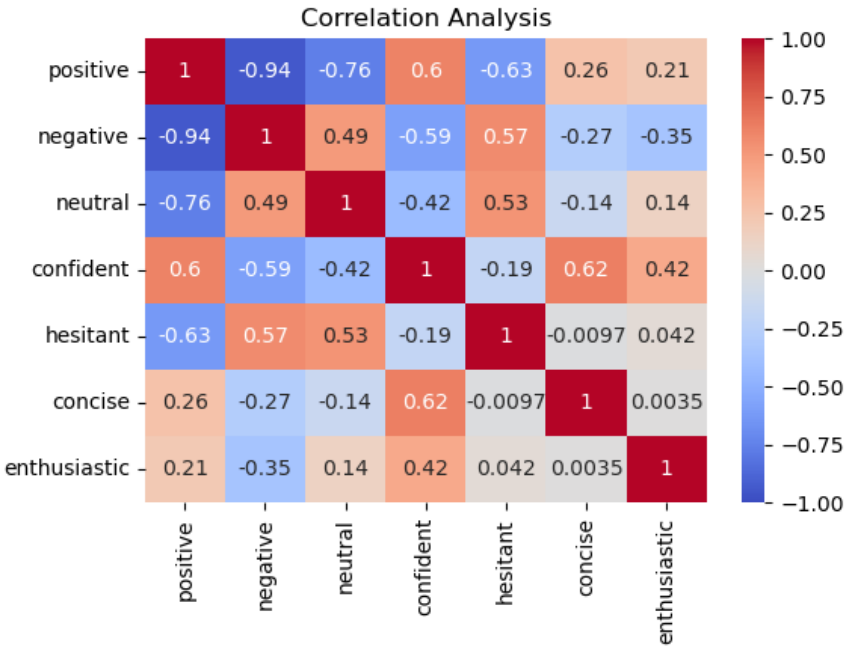
4. There are various valleys in hesitant graphs but the one which is worth noticing is when he talks about his ability to pay attention to detail, and his research work and patents, publication and the research award at IIT Kharagpur.

5. The person is mostly concise except when he repeats his ability to pay attention to detail.

6. The person sounds very enthusiastic when he talks about his hobbies of baking and traveling.

CORRELATION ANALYSIS:

Correlation analysis is basically a heat map which describes whether a particular emotion is related to another or not by specifying a value where strong correlations are generally (> 0.75 or < -0.75). This visual technique helps in spotting trends across different emotions and how the candidate traverses between them. For Person 1 the heatmap looks something like this.

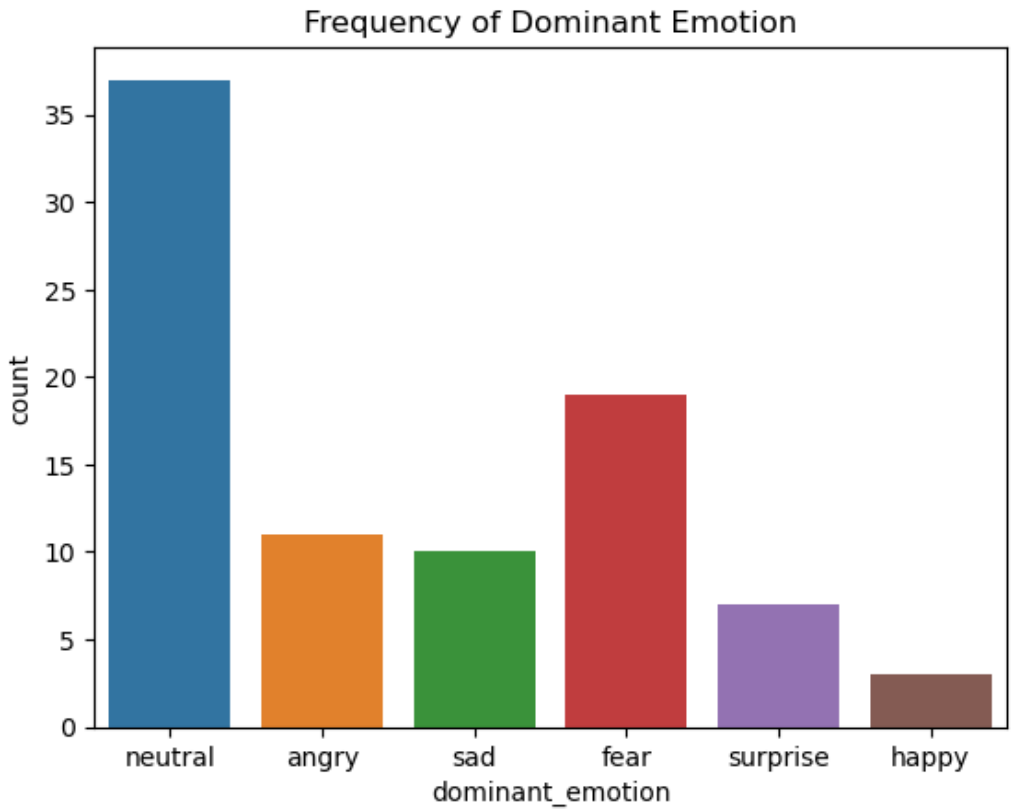


ANALYZING THE EMOTION DATASET:

In this section we analyze different emotions that are observed in the introduction video of the candidates and make certain inferences based on the outcomes.

DOMINANT EMOTION :

This analysis helps in understanding the dominant emotion that sustains throughout the movie of the candidate by a bar graph.



In this case the most dominant emotion is ‘Neutral’.

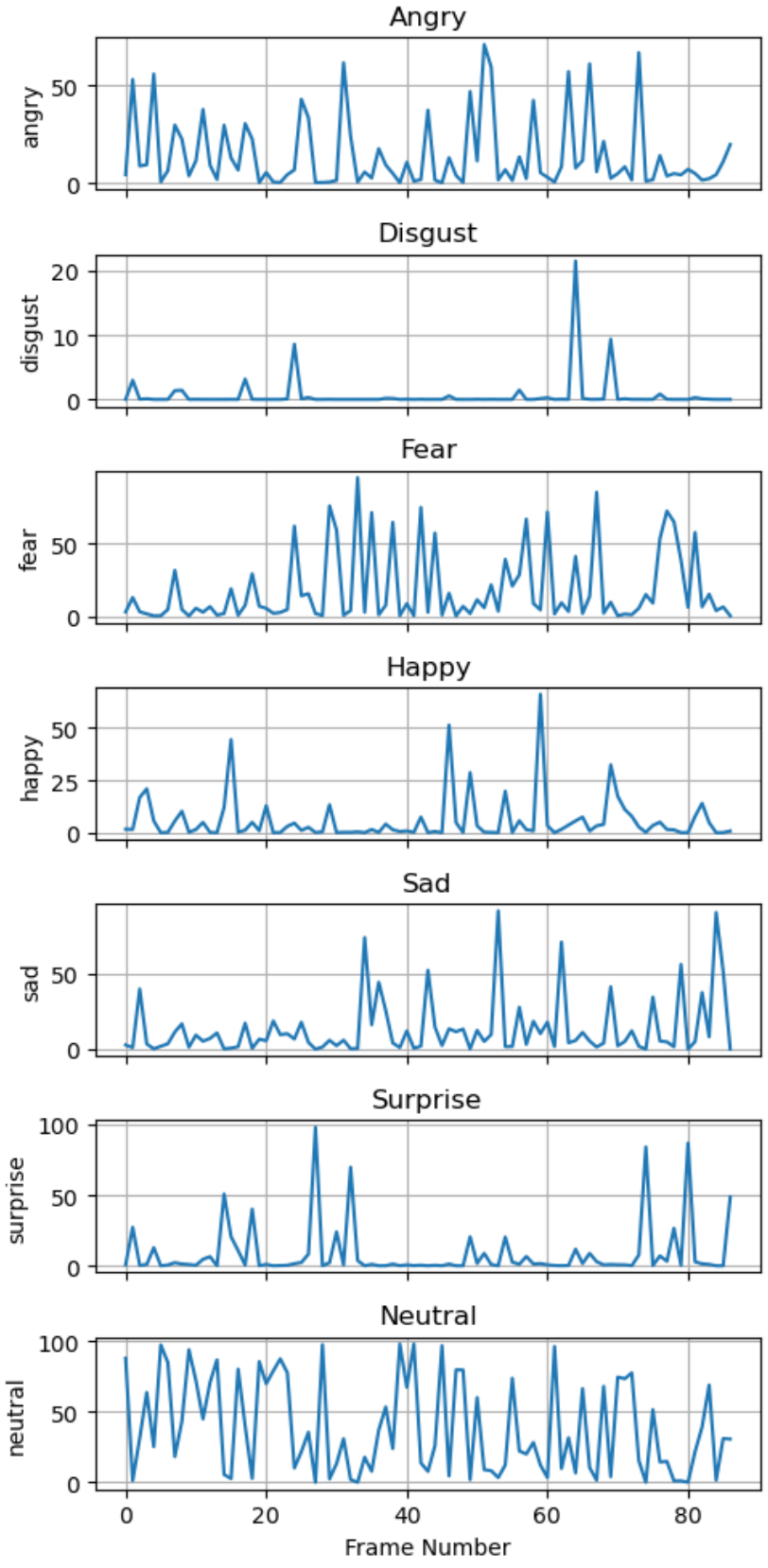
TREND ANALYSIS :

In Trend Analysis , like previously mentioned we analyze the trends of different emotions in the data. But this time we plot the graph of changing emotions with the frames of the introduction video and make necessary inferences.

P.S. The emotions mentioned here are emotions extracted from the video whereas the ones mentioned in speech analysis were extracted from the transcript text.

The trend analysis of Person 1 looks something like:

(ON THE NEXT PAGE)


Angry
angry
50
0
Disgust
disgust
20
10
0
Fear
fear
50
0
Happy
happy
50
25
0
Sad
sad
50
0
Surprise
surprise
100
50
0
Neutral
neutral
100
50
0
0
20
40
60
80
Frame Number

Here we make inferences based on the spikes observed in different graphs. Following were the inferences made for this case:

As observed from the graphs , there are constant spikes in emotions , 'neutral' , 'angry' , 'fear' , which indicate the speaker is providing factual or objective information with moments of stress, assertiveness, or passion about certain topics or experiences.

ANALYZING THE GAZE DATASET:

In this section we analyze the eye movements of the candidates through the video and make inferences about their confidence and preparedness by observing plots and charts.

EYE CONTACT ANALYSIS :

As the name suggests , we analyze what percentage of the time the candidate makes eye contact with the camera and based on that we make inferences.

For Person1 we get the following data by Python code:

Percentage of frames with eye contact: 62.50%

Percentage of frames with blinks: 0.00%

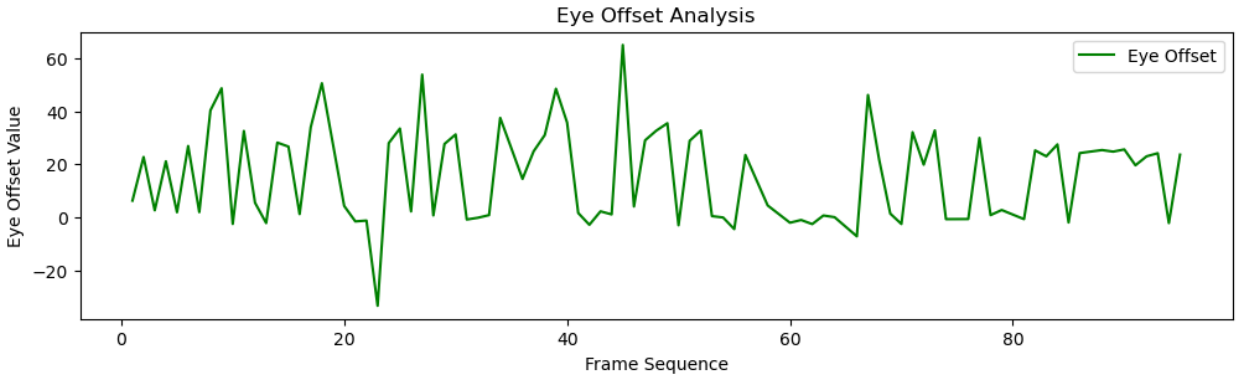
And we make the following inference from this information:

Given that the speaker maintains eye contact in 62.5% of the frames, it suggests a relatively high level of confidence and engagement with the audience. However, there are still moments where the speaker diverts their gaze, possibly to recollect thoughts or emphasize certain points.

EYE OFFSET ANALYSIS :

We analyze in this section how the candidate moves his eyes while speaking in the video by plotting a line graph of eye offset values against the frames of the video and from the graph we make inferences of whether the candidate looks at camera all the time or looks elsewhere while speaking.

For Person 1 the graph looks as follows:



We make the following inference from the graph:

The frequent peaks in the eye_offset graph indicate that the speaker's gaze deviates from the center regularly. This could be due to reasons like referencing notes, emphasizing certain points, or potential distractions. Such variations might also indicate moments of contemplation or trying to recall specific details or in some cases lack of confidence.

SUMMARY:

In the Jupyter notebooks of different candidates , I have also mentioned a summary of all the findings done from the above graphs. For the person1 the summary is:

The speaker mostly seems confident and positive but does get hesitant at times, especially when talking about past roles and skills. They make good eye contact for most of the talk, but there are times they look away, maybe thinking or just getting distracted. The emotion chart mostly shows a neutral vibe, so it's pretty chill overall. They get super excited when talking about hobbies like baking and traveling, though.