Question 4

Emotion	My Speech	Podcast Speech
Happy	Happy is characterized by the elevated mean pitch, which explains the excited and expressive richness typical of happiness emotion. The standard deviation(variability) in pitch is also high, which suggests a wide range of tonal expression. The mean intensity is also amongst the highest, while its standard deviation is moderate which shows consistent, dynamic energy that I tried to portray. The speaking rate is quick, aligning with animated articulation	Happy has a high mean pitch and the second-highest standard deviation in pitch, which indicates – wide range of expressive tonal variations that is often present in the joy state. The mean intensity is pretty high, combined with moderate s.d, indicating vibrant, dynamic vocal expression. The speaking rate is the fastest among all other podcast emotions, which reflects energetic, enthusiastic nature of happiness.
Angry	Anger has lower pitch than Happy, but still higher than Neutral. The pitch standard deviation is less than Happy, because of the narrower tonal range, due to more focused articulation. The mean intensity is slightly less than Happy but standard deviation is higher, which suggests bursts of voice to make accents and accentuate anger/hatred emotions	Angry has the highest mean pitch of all emotions, which indicates raised vocal tone that is part of aggressive, emphatic speech. It also has lower standard deviation in pitch compared to Happy, due to focused, less varied expression. The intensity has the second-highest mean because of strong vocal projection, when we are angry we raise our voices, and standard deviation is high due to volatility of anger.
Sad	Sad has a higher mean pitch compared to Neutral, with greater variability than Angry, however lower than Happy to indicate the feeling of sorrow, pensiveness. The mean intensity is lower than the emotion of Happy, because the voice is more softer compared to other emotions. The speaking rate is moderate, not rushed.	Sad emotion presents a mean pitch that's higher than Neutral but lower than Happy and Angry, aligning with subdued yet fluctuating pitch. The s.d in pitch is the lowest, suggesting a narrower range of tonal expression – emotions are more constrained. The mean intensity is close to the highest values and has the highest standard deviation, due to fluctuations in vocal force

Fearful	Afraid has the highest mean pitch compared to all other emotions, reflecting – heightened vocal pitch which is associated with “fear”, being flustered and in a state of shock. The feature of “pitch standard deviation” is also high which shows erratic, tense vocal patterns. The intensity mean and standard deviation are also overall on the higher side to portray the feeling of stress because of seeing “the cat” which could be interpreted as a fear of it could mean for the future/superstitions	Afraid has the lowest mean pitch which could reflect speech pattern subdued due to fear, but it has highest standard deviation – which could portray the sporadic peaks in pitch . The mean intensity is pretty high, and its standard deviation is the highest, again supporting the claim about unpredictable and sporadic vocal dynamics.
Surprised	Surprised emotion has a substantial mean pitch that's pretty high, however much lower than Afraid, because there is a spike in the beginning, the initial surprise and then it returns to the baseline. Pitch’s mean standard deviation is moderate, which indicates a brief but some variations in tone. The mean intensity is high, but also not the highest, which is supporting the previous statement of initial shocked reaction of seeing a cat unexpectedly that is not prolonged. The speaking rate is moderately fast.	Surprised has the second-highest mean pitch that reflects startled exclamations typical of surprise. It also has highest standard deviation in pitch, thus explaining a tonal variation. The mean intensity is high, and the standard deviation is moderate, indicating a marked but still somewhat controlled variation in loudness. The speaking rate is slightly slower than Happy possibly capturing the momentary hesitation
Disgusted	Disgusted has a mean pitch that's lower than Happy and Afraid, to portray the disdainful, more lower tone. The pitch variability is pretty high, compared to others, due to strong variation in tonal expressions because of the distaste and accentuation on things. The mean intensity is similar to Neutral, lower than Surprised and Afraid, and its variability is high which suggests sporadic emphasis on observations which were in this case “ cats”.	Disgusted emotion shows a mean pitch lower than Happy and Surprised but higher than Afraid, which could reflect tone of aversion / disdain. The s.d in pitch is moderate, mean intensity is comparable to Neutral with low standard deviation showing consistent but reserved vocal force. The speaking rate is the fastest, surpassing even Happy, which may represent the quick dismissal

Neutral	Neutral is the baselines with the lowest mean pitch due to calm, steady tone that also has low variability – there are no spikes/accents, everything is stable. The mean intensity is moderate, and standard deviation is also lowest, which again indicates the balanced, controlled and emotionless vocal delivery – there is no positive/or negative connotations after seeing a cat on the street	Neutral has the lowest mean pitch, also being the baseline for comparison. Its s.d in pitch is relatively low, showing steady and even tone. The mean intensity is not the lowest but shows a high s.d, suggesting more variation in loudness. The speaking rate is moderate, faster than Sad but lower than happy and disgusted.

Question5 -

1. What are some similarities and differences between the features from the two datasets?
Differences - The datasets present notable disparities in the acoustic characteristics of Angry and 'Afraid' expressions. In the podcast dataset, Angry is characterized by significantly elevated pitch, suggesting a portrayal of anger that is more intense or dramatic, whereas in contrast, my speech depicts anger with a subdued pitch, indicative of a more restrained emotional display – which may reflect different way people express anger. For Afraid, the personal dataset records the highest mean pitch, aligning with the conventional association of fear with a tense, high-pitched voice. However, podcast dataset presents it with the lowest mean pitch, albeit with a high standard deviation, indicating subdued tones punctuated by moments of heightened pitch, reflective of sporadic intense fear. The representation of Sad also diverges – the podcast dataset portrays it with greater intensity, to emphasize the emotion, whereas in the personal dataset, sadness is conveyed through a softer tone.
Similarity – Both datasets have a high mean pitch and intensity which underscores a universal inclination towards expressing happiness with vibrant and energetic vocal tones, being excited and emotional. Similarly, Disgusted is represented in both with a lower pitch relative to Happy and Afraid, suggesting shared vocal expression of aversion or rejection, not accepting and being disgusted. Neutral alsos serves as a common baseline, characterized by the lowest pitch, with calm and stable emotional state without heightened emotional expressions.

2. Which of the datasets would be more useful for emotion recognition applications? Why?
Podcast speech because it may offer higher value due to its encompassing range of vocal expressions and ambient conditions. This variety can facilitate development of more adaptive models capable of interpreting — a more nuanced spectrum of human emotions across diverse real-world settings. Having different speakers(different genders) and situational contexts enhances the potential to train systems/models in recognizing a broad array of emotional expressions – this would lead to more robust applications.

3. Which of these datasets would be easier for an emotion recognition system to classify?
My speech dataset,due to its consistency in speaker identity and possibly uniform recording conditions and settings, thus making it a simpler task for emotion recognition. It becomes more standardized, the homogeneous nature ensures lower variability and leads to more straightforward and easy ways for recognition systems. It would be easier to classify the emotions made by one person and also the length of the sentences I used were shorter which also makes it easier to classify.

4. What other features would be useful for emotion recognition?
Additional features that could be useful are Prosody because the differences and variations in speech rhythm, stress, and intonation are crucial for understanding the differences in emotions. Additionally, Mel-Frequency Cepstral Coefficients (MFCCs, could be used to capture the sound's power spectrum over time, providing important and critical information, including the time change aspect for voice recognition tasks. And finally non-Verbal Cues, specifically sounds such as sighs, laughter, or crying that are usually edited out of the recordings, could serve as potent indicators of emotional states and lead to more nuanced/detailed analysis.