

Initial thoughts:

- Could potentially use RegEx to pull out the data
- May be worth segmenting the text if it was a larger transcript? Could possibly be split at the new lines
- Need to allow for different formats
 - In the transcript the keys are not all in the same format case wise ("NHS Number", "NHS NUmber") so need to account for variation in casing - possibly normalise text
 - Keys also do not always have a colon
 - Date is provided in multiple formats (2nd of May 1986, 03 01 01) so need to make sure all cases are captured
 - Fields can be empty, but not all fields can be empty at the same time - actually, all fields have an NHS number
 - NHS Number does not seem to have a standard configuration it fits - some have letters, and they are of varying lengths
- Is there an inconsistency between the transcript and output? NHS number 999 seems to state age 48 in transcript, but is 38 in output - indicates 38 was calculated through DOB rather than from transcript. Likely the same for Micheal who is stated to be 24 as age isn't explicitly labelled

Blockers:

- I haven't used React before
- What if patient has name with middle name listed? Or title?
- Not necessarily relevant for this task with test data, but if expanded into full system would possibly need to take into consideration things like GDPR and HIPAA as these are patient records

Retrospective after completing:

- I would like to make the react front end look nicer
- Issues with regex - doesn't currently detect date of birth for one of the records as current regex looks for records in certain order (but I ran out of time to fix this)
- Would love better security around database login info
- Want to set up an input/output page so you can view all data in the database
- Currently spamming the submit button will keep adding data to the database
- Regex misses out on one unlabelled name
- Could possibly be improved processing had I segmented the text - could use that to make processing better for if really long transcripts were given
- I did wonder when programming if there would be a way for better processing NHS numbers from audio to transcript as I don't believe NHS numbers have letters in them, and there were occasions where I could see numbers possibly being mistaken for letters (4t could sound like 40)
- Duplication checking for the database