**M.Sc. Thesis in-depth interview-20230809\_104248-Meeting Recording**

0:03  
Yes, I think the recording is on Mary Okay, yes.

0:07  
So Mary, I will, I will just go to the agenda 1st and then I will, I will share the screen and I think I shared it to you by e-mail.

0:34  
Yeah, it's on.

0:35  
It's in the meeting.

0:37  
Yeah request.

0:46  
So going by the agenda like first three points is something like I will explain the overall design and what is the primary research and what is the sampling I'm using just to set a context to the questions.

0:59  
OK, so this this meeting is like as part of the thesis which I'm which I'm can you see my screen?

1:17  
I can yeah.

1:17  
Yeah.

1:18  
So this is this is as part of my MSC course in data analytics.

1:23  
So as part of the course I'm supposed to complete thesis.

1:26  
So I have identified A thesis which is appropriate with my work so that I don't have to.

1:32  
I don't have to deviate too much between work and this one just to manage time well.

1:37  
So the the thesis topic I've chosen is with the title Predicting the screening colonoscopy numbers across Ireland using Machine learning algorithm.

1:48  
So what are the objectives is like there are three.

1:51  
I have divided the objectives into three.

1:53  
The first objective is the statistical objective like just to identify if there is any variations between age and gender.

2:07  
The this was like the the the idea of doing this is like if if there is substantial variation with age or say gender and the the colonoscopy units can plan better like we are planning to extend the age by one age at a time and then like if you see that there is a huge variation and basically the capacity capacity planning in the colonoscopy units can be done better.

2:33  
So that is the first objective.

2:35  
The second objective is to see the correlation between different features in the colonoscopy.

2:42  
So this is where the entire primary research of in depth interview comes into place like as of like what we what I see in the data the features which which could potentially impact colonoscopy are the gender and age.

2:58  
Other than that like I I would like to know like if there is anything more which which I need to consider for predictions.

3:05  
The third one is the third one is to is to actually build a machine learning model and machine learning model will be built based on the existing colonoscopy data and it will be the the data will be extrapolated using the census data for the younger age groups like because we we have the data only from the age groups of 60 to 69 as of now.

3:33  
I will use the census data identify the percentage of population who has undergone colonoscopy and extrapolated to create the data set.

3:41  
And using that data set I will try to predict it for the future future years.

3:46  
So these are the three three main objectives and yeah I've I I missed to mention something I will I will also be doing an hypothesis testing to see to actually see if the colonoscopy counts in males are less compared to females.

4:04  
That is again depending on that like if the the null hypothesis will be like if in the depth interview I I come to know like like I I if everyone in the dipped interview suggests that they are actually the same then the null hypothesis will say male and female are actually the same.

4:23  
So it it it, it varies that way.

4:28  
So these are the objectives.

4:31  
And then moving on to the sampling strategy like as as part of the research I'll have to do sampling strategy.

4:39  
So there is no the like, there is no much significance of the sample amount because the population we have is not substantially high.

4:49  
We can do all the study based on this population itself.

4:53  
But then what I have done was when when you say sampling strategy, I have identified 2 populations like I have divided this data set into two populations male and female and the data between 60 to 69.

5:10  
They are actually the sample and the date the the extrapolated data from 55 to 55 to 5025 to 69 that is the population that is the whole population.

5:24  
So the average will be calculated for the sample which is selected of 60 to 69 and using this average I will be doing my statistical analysis for the bigger bigger data set.

5:38  
So that is again hypothesis testing and all that, right?

5:41  
It will be, it will this is this is significant for the hypothesis testing that is of the sampling and then the primary primary research.

5:53  
I have selected in that interview as the primary, primary research because I could have gone for something like in experimentation kind of primary research.

6:03  
But I could not go there because theoretically it did not satisfy the criteria for experimentation because I wanted something with where the colonoscopy numbers varied with the time, such kind of thing.

6:18  
So I did not find anything like that.

6:20  
I couldn't like there, there wasn't any significant changes with season or months or something.

6:26  
It was all like more or less the same.

6:29  
Yeah.

6:32  
Yeah.

6:32  
So these are the, these are the three first three points I had to discuss, right.

6:40  
It's very interesting.

6:42  
Yeah.

6:42  
So these are the, this part of the agenda.

6:44  
I have discussed the first three this one.

6:46  
So are you OK to go to the questions now?

6:49  
I am, yeah yes.

6:52  
So like I had just framed these questions like one is based on something which is which will align to my objectives and also based on my in depth interview with a couple of others.

7:08  
So based on based on that I have framed this interview questions.

7:14  
Maybe one thing is like you can give the answer as generic or detail as you wish but okay.

7:19  
Please make sure that there is no sensitive information or something which has not gone public.

7:25  
Not yeah mentioned in this meeting because this is a recorded meeting and it'll go go to the college.

7:30  
Okay, I'm gonna crash.

7:32  
Thank you.

7:33  
Yeah, Okay.

7:36  
So first my first question is like when in in bubble screen program participation, do we actually see hesitancy at a younger age like I know that we have currently between 60 to 69, so we had two groups 60 to 65 and 65 to 69.

7:57  
So as far as I have seen in the program reports, I I saw that in the first program report there was a variation 60 to 65 had significantly high numbers, but in the second program report it was more or less the same.

8:13  
But as far as your experience is concerned, like do you, do you actually foresee any, any lesser or more people to participate in the program?

8:27  
So I suppose as you say, Rakesh, looking at the round reports we would monitor, we would observe the participation rates amongst the different age groups.

8:37  
And as you say in most recent round reports, there doesn't seem to be any significant difference the age range for the round reports that you're referring to 60 to 69 years old.

8:50  
And as part of the National Cancer strategy, the program is planning towards age extension and that will the program will extend over a phased basis to age 55 to 74.

9:00  
So I suppose as we do start age extent, we will look at both the upper and lower and age groups to monitor the uptake and and to monitor participation in the program to see if there's any any differences in other measures.

9:15  
I think that the program is aware of to try and ensure that UM to maximize participation is to to have targeted communication and to to educate the public in terms of screening and and what screening is.

9:29  
And so I think it's one that we will continue to monitor, but in the current age range as you say there doesn't seem to be a a major significant difference Okay, Okay and my.

9:47  
The next question to you will be like with regards to men and women undergoing the screening, like are they equally probable of participating the screening Like is there like I think in the in the earlier study somewhere by DCU by Nick Clark, I could see that there was, there was participation rate in men actually less compared to women.

10:16  
Yeah.

10:16  
So, yeah.

10:17  
So like what do you think about the probability when you go for the younger age and well, I suppose just first of all go back to the study that you mentioned it was, it has been a feature where men are.

10:31  
Research has indicated that men are more likely to participate with female influence in some cases.

10:39  
I suppose in terms of the younger age group, it's really difficult to say it's it's something again that we will closely monitor.

10:50  
I think again going back to having a very good targeted, targeted communication strategy will play a role within that.

11:01  
So it's it's it's one that we're going to watch closely.

11:04  
But I suppose until we we do that age extension piece, it's a very difficult question to answer.

11:11  
We've no reason to believe that it will be that there will be any major change.

11:16  
But like that it's something we monitor and work to try and ensure that we can try and keep the participation numbers amongst men and women as as high as possible.

11:27  
Okay, okay like was there any like apart from Ireland like some other country would have done that Like is there any reference or something which you can give me like which is country I I'll try to get the program report from them to understand it better.

11:44  
I think yeah there there there is.

11:46  
The program does a lot of research we do a lot of behavioral science research.

11:50  
We work we we do a lot of research in terms of participation and what the motivators are in terms of international research.

11:58  
I'm sure it's it's it's very much the same um.

12:01  
I suggest that you could find numerous studies out there on it.

12:06  
Um, I I wouldn't have any to hand at the moment, but I'm sure there are studies that have looked at the what influences participation.

12:16  
Um, quite possibly they would look at at that in both males and females.

12:20  
But um, as I say, you know it, I would maybe suggest you look at the research done in DCU and look at their references to see what the international experience was even that might be helpful.

12:32  
Yes, I will.

12:33  
Thank you.

12:33  
Thank you, Mehdi.

12:37  
OK.

12:37  
So moving on to my next question, are there any other measure or feature that I need to consider?

12:47  
Like for now, I think I have considered only only the age and gender.

12:51  
But apart from that, is there anything, anything else which I will need to consider to make accurate predictions?

12:58  
Is this accurate predictions in terms of the colonoscopy numbers that are required?

13:05  
Yes.

13:05  
Accurate future colonoscopy numbers?

13:07  
Yeah.

13:08  
So I suppose when you're looking and you're making your predictions, you're making your predictions based on the census data.

13:16  
But the program's data source is the Department of Social Protection, so there will be variation between the Department of Social Protection data and the census data.

13:26  
So that is certainly I think something that you would need to consider um also I suppose you know it is the accuracy of the data, so from the Department of Social Protection.

13:37  
So we are depend we we will invite people based on the addresses provided to us by them.

13:44  
But if that is inaccurate and or hasn't been updated recently um then those those letters will um be returned to the program.

13:53  
So I suppose part of the our what we're doing within the program is trying to ensure that our register is kept as accurate and as up to date as possible.

14:02  
So within the team we have dedicated resources who would manage that piece for us and you know but like that there will there will be variation I think between both, both data sources.

14:15  
Okay, okay, okay.

14:25  
Moving on the next question, like are there any research done in the past with regards to the population extension if S what are the features considered for the study?

14:35  
Like I think you just mentioned there are numerous studies available but is there anything like for example anything from NHS or anything from Scotland or something of that sort like yeah I suppose for first of all it's supposed to speak about balance screen.

14:54  
So for balance screen as part of the National Cancer strategy we have to we're required to extend the age range from 55 to 74.

15:05  
Now the program regularly reviews evidence for screening of of younger age groups UM and it may be while the current plan is to extend to the to the UM range of 55 the the downward range would be 55 that may be younger UM and it could go to 50.

15:24  
So UM.

15:25  
For example in in Scotland UM they begin screening at 50 and then it's 50 to 5 to 60 in other parts of England and I think it's 55 in Northern Ireland.

15:36  
So there there is variation amongst amongst the UM, the different areas, UM I suppose quite for in order for the program to extend our age range we would need to seek, that is true the National Screening Advisory committee, so the National Screening Advisory Committee on behalf of the program have been looking at this and they have requested a specialist team in Hickwit to look for evidence for the expansion of the age range and from 50 to 74.

16:08  
Okay, okay, okay, yes.

16:15  
So I think I will what I will do is I will try to find any materials or any, any literatures wherein like you mentioned about Northern Ireland and Scotland.

16:26  
So I like they could have some program reports which which which will show like how the participation is for different age groups.

16:34  
So yeah that may be detailed in in their their reports as they say it's it's an ongoing piece where we we look at it's based on I suppose best practice and and what the international evidence is showing us and the the program will as I say extend from 55 to 74.

16:53  
But but anything a greater Adrian extension is something that would have to be reviewed and is is ongoing.

17:03  
Yes, okay, okay, got it, got it.

17:07  
Yeah.

17:09  
And I think my next question is what are the practical challenges you see while predicting the colonoscopy numbers based on the census data.

17:20  
So again, I I suppose the practical challenges are that the census isn't our data source UM.

17:26  
And as I mentioned already, our data source, UM is the Department of Social Protection.

17:30  
So it's UM, it's trying to ensure that the data we have is up to date.

17:37  
It's also trying to ensure I suppose that you know those that are included in the census, UM are on our register.

17:44  
So.

17:45  
So we actively encourage people through I suppose our communication strategy to check the register to try and to register for the program um.

17:56  
So I suppose it it's it is that it's it's not our data source And and I think that would be one of the challenges um when we would look at at predicting colonoscopies, we would look at the the our own data that's available to us.

18:11  
Um.

18:12  
So we work as I said while ago, we have within the program, we have dedicated resources that work on our register continuously to try and ensure that it's accurate and that there's no duplications.

18:25  
All of these pieces um, it's very complex process, but it is something that is is something that is very important within the program because it does impact our participation rate.

18:36  
UM, they if we if our data isn't accurate.

18:40  
So it's something that we are very, very aware of and something that we work very hard on trying to keep as accurate as we possibly can.

18:49  
Okay Beri, just as a follow up question for that, right, like apart from the people in the register like for example, we have something like Men's Health Week for example.

19:02  
And yeah, we, there are social media information going on for the Men's Health Week and then anybody calls up and tries to register for the program, right.

19:12  
They are not in the register, right.

19:14  
So such kind of things like for example we have, we have all this the immigrants or the refugees based on say from the Ukraine or something like yeah, Ukrainian refugees they call up and they register for the program.

19:33  
So such people will not be uh, there currently, but when they call up only then they will come to the register, correct?

19:42  
Um, well, if they've had interaction with the Department of Social Protection and they're within the eligible age range then we will get that information.

19:51  
We will get their their their contact details in terms of of of writing to them, yes.

19:57  
But what you say is yes, like we for example in April, it's bowel Cancer awareness Month and we would do a lot of promoting of the program and we would encourage people to check that they're on the register um and then to try and make it as accessible as possible.

20:12  
We have a lot of our leaflets and information available in in various languages which can be accessed to our website.

20:19  
So we want people to be able to participate.

20:22  
We want to make.

20:23  
Access to the program as as easy for people as possible.

20:29  
In terms of of talking about individual groups, it can be difficult to say because I'm not sure and you might be able to answer this for Kesh.

20:38  
Has this been looked at your within your department?

20:42  
Have you looked at the census versus what's on the registered and found evidence that there are large groups of people that haven't been that aren't on the register.

20:57  
There are, there are no, I mean the level of data which is available from sensors is not like we we I don't think we have very detailed level of data.

21:09  
But overall in the percentage there is a difference, there is, there is I think at least for the lower age range there is around 15 to 20% difference.

21:18  
Yeah.

21:18  
Between what is there in the sensors and what is there in the yeah register that that is that I think though 20% is quite high, some difference is expected because the whatever is there in the register is like we we will have, we will have many people at that point of time on that day particularly because it is a weekend and that it's in April.

21:44  
So somebody would have people would have come to airline for a visit or something.

21:47  
So, yeah, yeah, yeah.

21:49  
And I think that's you know you will have variations but we receive regular files.

21:54  
So as I say if there has been interaction we and there for people within the eligible age range we would we would I suppose we would hope that we would receive that data.

22:06  
But then to try and I suppose make increased awareness of the program and and to try and encourage people to register and participate, we would have communication, UM communication that would target specific groups, UM and and general communication and and an overall strategy within the program.

22:31  
And then as I say, we would also try and make materials available so that people would be able to understand what the program is and what screening is and it would be able to answer any of their questions in terms of it.

22:43  
So that's work that has been ongoing in the program I suppose.

22:46  
So that we can increase the accessibility for people and you know, ensure that that people can participate if if it is something that they would wish to do.

23:00  
Yeah, Yeah.

23:01  
Yeah.

23:01  
So I think then in that in that way, like if you see it more or less, the census data will be quite accurate only because it might not be 100% though we are considering this one because because I'm I'm taking the percentage of people who has undergone colonoscopy in that census.

23:23  
Yeah.

23:24  
That percentage should not ideally vary much.

23:27  
Yeah.

23:28  
Yeah, OK.

23:31  
It's it's difficult as I say because you've two different data sources I, I I wouldn't be, I wouldn't be able to say that they're the same it's, it's it's two different data sources.

23:41  
So I I think that might be a challenge you may have, but I suppose from the program level we work to try and ensure that the population within the age range you know our our registering and are aware of the program so that we can maximize our our participation but also that our data is accurate.

24:04  
And we constantly look at measures to see is there more we can do to try and and make the program more accessible to the public.

24:12  
Um.

24:12  
And that's where we look at translation.

24:15  
Um and you know I suppose just different using our website and using even social media and things like that as as a program you know matures and things.

24:26  
We're we're trying to look at different ways of communicating with people and and different ways to that will will see the eligible that will I suppose be visible to the eligible population.

24:41  
Yeah.

24:41  
Yeah.

24:42  
Yes.

24:44  
I think.

24:44  
Yeah, correct.

24:45  
What I think, I think what is it, it makes sense actually but only thing is like it when you do the predictions because I'm doing through the census.

24:55  
I'm just wondering like if that will be correct.

24:58  
It should be I think right because predictions like predictions are for the capacity planning and you always plan for some at higher level.

25:07  
Well, I I think it's useful really it will be very useful because until you know and I suppose we it does give us a an ability to see what the demand in the future will be.

25:21  
Um you know for as we as we do implement phased age extension um and the census is something that we always use as a benchmark as well against for our own, our own register data.

25:37  
So it's trying to see if there are variations and look at why there are variations and you know if, if, if that would suggest that we need our our own information maybe needs um further clean up or if it's you know we look at it to see where we've made improvements.

25:54  
Um.

25:55  
So from the work we're doing we would often look to see compare that to the census to see um if it if it is coming across as being more accurate.

26:04  
Okay, just just as a follow up question on this, right, like is there any specific reason why why we had gone to the Department of Social Protection data and not the census data for for building the register.

26:18  
It's a the census is only every is it every four to six years.

26:26  
Well we would get regular data imports.

26:29  
So I suppose as you said yourself the the census is done on a night at a point in time while we would use we our data source we we get that frequently.

26:41  
So it's that's probably part of the reason I would imagine when the program was established, that was the data source that was that was identified Okay, okay, no problem.

26:59  
I think I'm OK with questions.

27:01  
Yeah.

27:01  
One, one last question based on the discussion with the Terrace yesterday.

27:08  
So she was mentioning about the fit scores, the fit score, what is that cut off, fit score, cut off being changed somewhere in 2014.

27:21  
And she had mentioned prior to that the holoscopy numbers were very, very high.

27:26  
If we use that higher colonoscopy numbers, we will not be able to actually do an accurate prediction.

27:32  
There will be some miss like misalignment in the overall colonoscopy numbers.

27:37  
Yeah.

27:37  
So yeah, So is there any specific date or specific month or quarter right, like from which I can start the cutoff like currently I'm using the data from 2014.

27:49  
So so you're using the data from 2014 to sorry to determine what I I what I mean I the colonoscopy numbers I'm taking from 20/20/14 to 20/22 December.

28:04  
Yeah.

28:05  
And then I'm I'm my all my predictions are based on this data.

28:10  
OK.

28:11  
But are you looking at the overall positivity score like the most recent?

28:16  
Because that would be based on the, the fist cutoff range.

28:23  
Yeah.

28:24  
No, what I'm saying is like if I go for say some date in 2013, the colonoscopy numbers will actually be higher, at least the percentage of people.

28:34  
Yeah.

28:34  
Or another colonoscopy will be higher because the FIT score was low threshold.

28:40  
Yeah.

28:40  
Was low, Yeah.

28:41  
So now we have increased the threshold.

28:43  
Yeah.

28:43  
So because of that, because of that it is, it is like it'll be different than how it used to be.

28:52  
That's it.

28:54  
But for future predictions, yeah, if you look at the most recent published round reports, I would suggest using that positivity score because OK, that's what the they were based on the current threshold.

29:07  
OK, OK.

29:08  
OK.

29:09  
OK.

29:10  
I will, yeah, I will figure that out.

29:12  
If I see substantial high, then I think I will remove that from the research altogether.

29:17  
OK, Yeah, yeah, yeah, yeah.

29:22  
I think that's all I had.

29:26  
And one more thing like the the fit score, the fit score threshold value, right.

29:30  
It is, it is same for both the genders, correct.

29:33  
Yes.

29:34  
Yeah.

29:34  
Yeah.

29:34  
Yeah.

29:34  
Yeah.

29:35  
OK, it is.

29:36  
Yes.

29:37  
OK.

29:38  
So it's it's the same for both genders across the eligible age range.

29:44  
OK, OK, OK.

29:45  
And are we looking at reducing the, this one, I mean increasing the threshold value when you go for younger age range, No, there's no immediate plans.

29:57  
I think that's again that's a very, very big program decision and that would involve I suppose are a lot of from our input from our clinical advisory group from public health putting forward recommendations as I said earlier to the National Screening Advisory Committee.

30:14  
So there there's as it stands which the plans to implement age extension.

30:19  
I believe that the, the FIT score will remain the same initially anyway, but who knows down the line.

30:27  
Yeah, yeah.

30:28  
Yeah.

30:28  
No, the reason I came up with that question is because my my studies are purely based on the colonoscopy numbers and census data is used to just to extrapolate the data for the younger age group Okay.

30:42  
So I have not considered uptake numbers.

30:46  
I have not considered fit positivity numbers.

30:50  
Yeah I have not considered because if the fit cutoff numbers and all those things are the same across the age range then it makes sense to simply use the colonoscopy numbers.

30:58  
Otherwise I should see like I should consider this percentage variations also in my research.

31:08  
Yes, I think Mary, I think this is all I had for for this in depth interview.

31:16  
So okay.

31:17  
I yeah, I will, I will now thanks a lot for your time Mary.

31:20  
And I will, I will now proceed with the next steps.

31:23  
I will try to make a report and I'll keep you posted on my on the outcomes of this research and all that.

31:30  
Great.

31:30  
Well, the best of luck in your your research and in your thesis for question.

31:34  
Thank you.

31:35  
Thank you, Meri.

31:36  
OK, take care.

31:37  
Thank you.

31:37  
Thank you.

31:38  
Thank you.

31:38  
Bye, bye, bye, bye.