**GYE DA\_1\_COHORT 2025-20250317\_091926-Meeting Recording**

March 17, 2025, 9:19AM

1h 11m 42s

 **Emmanuel Kofi Essel** started transcription

 **Emmanuel Kofi Essel** 0:05  
Lydia, who led the session this morning.  
Was asking us to decide what AI is going to. What we will ask AI to do for us depending on whether I want it to clean the data or whatever it is. Because this morning session is practically the same thing that we have done. I mean.  
Lydia, do you? Was it revealed to you in a dream or something?

 **Lydia Kwao** 0:33  
No.

 **Emmanuel Kofi Essel** 0:35  
I see.  
Anyways, so this morning we are having.  
Something that was titled AI Challenge, we are looking at ethics and.  
Ethics in AI driven data analysis.  
We have.  
We have done AIA couple of weeks ago. We have we understood how you could use AI to help our work.  
Data cleaning looking for.  
Problems in our data set. Asking AI to guide you as to how to clean your data and all that, and even to the extent of going ahead to tell AI to do the cleaning for you.  
Now, as great as that may sound.  
There is another principle that we may have to look at.  
It is the principle of.  
Biasis.  
Right is a principle of biases.  
How many of us?  
Have been in a situation where we think something done wasn't fair, it was bias.  
This one I know everyone of us have done some before, so I want two or three people to share.  
A situation where you think that the decision that was taken wasn't fair. It was bias. Someone someone was preferred over the other.  
Yes, anybody to help us.  
It can be esion de escalate the one it went into it entered into your heart and brought out your soul.  
And if it wasn't for God, you hold a person and tear him or her apart, OK?  
Let's start with Jeremiah.  
This week you are the first person responding to a question in class.  
Oh, that is a milestone.  
You should put it on in, yeah.

 **Jeremiah Elorm Aleawobu** 2:35  
Think you said so?  
This one is very practical.  
It's. It will happen yesterday.  
I'm I'm talking in terms of football.

 **Emmanuel Kofi Essel** 2:42  
Wow.

 **Jeremiah Elorm Aleawobu** 2:45  
So yesterday, Basa had a match and it was against Atelescu Madrid and yesterday the score Debo, which was a handball, but they didn't check by and still give up. They go to Athletico Madrid, which was very, very unusual and very unfair.

 **Emmanuel Kofi Essel** 3:08  
Wow. Wow.  
So let's go. Madrics called the bull, and it was supposed to have been called out for a foul.  
That's what you are saying.

 **Jeremiah Elorm Aleawobu** 3:25  
Yes, yes.

 **Emmanuel Kofi Essel** 3:26  
Right, right.

 **Jeremiah Elorm Aleawobu** 3:27  
It was a handball.

 **Emmanuel Kofi Essel** 3:29  
Oh. Oh, OK, OK.  
For football, I guess. And for Barcelona fans, it didn't really change anything. Anyway. At the end of the day, the people took like four goals.  
Diego Simone J4. They say they got to come to Cote, a Mani managers and so yeah, that was what happened. Great.  
Wilhemina.

 **Wilhemina Mensah-Doe** 3:58  
Good job.

 **Emmanuel Kofi Essel** 4:00  
Yeah.

 **Wilhemina Mensah-Doe** 4:01  
Ah yes, mine.  
It's it's when, when when I was a kid.  
I I my sister broke things in the House and I was beating 'cause. My mommy said I'm the eldestone.

 **Emmanuel Kofi Essel** 4:08  
Oh.

 **Wilhemina Mensah-Doe** 4:15  
I should have.  
I should have taken care of the pain.  
Meanwhile, both of us are politics.

 **Emmanuel Kofi Essel** 4:21  
Exactly. These are stories that I want to hear.  
Let's let's give it up for.  
Some of you you think it didn't happen to all of us?  
It happened. Yes, your mother will beat you for no reason.  
Like there's this beating that I received from my mother.  
So now I do understand why I got that, but she's my mother. So the mommy is always right.  
This biases some people are preferred.  
I want to come to the world of data analysis.  
There are situations where if we are not careful enough as data analyst, we are going to be bias.  
Sometimes we have a good intention.  
But.  
Because we have not done proper checks, our results or the analysis or the insights we glean from the data we are working with, it's going to be a little bit tilted towards somebody, OK.  
So that is what this morning we are going to look at.  
We will get to the objectives very soon.  
I have like 60 minutes to do this.  
We are ready 10 minutes in.  
So let's look at this situation.  
Imagine you are choosing a gift for your friend.  
I mean for your friend, you know, for that friend.  
Let me qualify it for for.  
And instead of considering their preferences, you base your choice on what you personally like.  
You choose from your book.  
You choose a book from your favor.  
Ite.  
Samsung, Samsung, assuming they will enjoy it too, even though they prefer something entirely different.  
Now, how many of us have received gifts that the people who brought the gifts thought we are going to be excited by? We were like.  
OK.  
Thank you, but no thanks.  
Have any of us have received a gift like that before?  
Thanks, but no thanks.  
I mean, it could have been better.  
Could have been better.  
What? What is this for? These people today thought they have done ghost job like they sense their son.  
They they are only begotten son to come and die for you. So immediately they come. You have to worship them and you know but.  
I think came and were like, OK, OK, OK.  
Well, thanks, but no thanks.  
Now, what do you think?  
The person or what do you think this person didn't do right?  
Yes, anyone.  
This morning I want my very own Amsterdam to talk.  
Amsterdam. What do you think the person didn't do right?

 **Monica Adom** 7:21  
Hello.

 **Emmanuel Kofi Essel** 7:22  
Yep.

 **Monica Adom** 7:23  
Hello please can you repeatulations.

 **Emmanuel Kofi Essel** 7:28  
I'm asking that somebody wants to gift you. You and the person has considered what he or she likes, right?  
And then bought a thing for you.  
Even though what you wanted was entirely different, but they have brought it.  
What do you think the person should have done right?

 **Monica Adom** 7:52  
Should have asked me what I wanted.

 **Emmanuel Kofi Essel** 7:56  
Should I have asked?  
Exactly. Thank you very much.  
So Monica, when he tells he's buying the 10 and he asks you, what do you want? You can also say I want this for some of you.  
Some of you.  
Some of you.  
There's this phrase popular phrase that you people use.  
I don't know.  
And Chairman will also you know?  
It is too early in the morning. Let me continue.  
It is too early.  
So what was wrong in the decision?  
That was what Monica said.  
The person didn't.  
I am tempted to use he, but knowing Amsterdam.  
The person didn't consider right.  
The person didn't consider.  
Money case.  
Preference. He or she could have gone ahead to ask, oh, I want to get something for you.  
What would you, at least Monica could have given options.  
I don't know if Monica really would have given options, but they'll go ahead and tell you you buy anything, anything but you buy the anything to and anything is not anything anyway.  
This is an example of a bias and some of sometimes it's going to happen to you in your data, right? You, you you have a lot of chances that some of the data you are you are using.  
Are advantaging some people it it has?  
It is a certain end, and if you don't make sure this data is fair, you are going to make another mistake.  
Now this one, let's come to the the first one was for the girls, right?  
You know, the girls like that.  
I don't know, but when they bring it, you still don't appreciate it.  
I don't know.  
I'm not speaking from experience, by the way.  
I'm speaking from examples that I learned, so I've also done machine learning and I have gathered the examples.  
Now imagine you are selecting A-Team captain for a school football match and you based your decision on past game performances.  
But you only recall performances from matches where it did not rain. So because it did not rain, there was this player that played excellent.  
Now.  
You are looking at this person from the point of view that, oh, this guy is a good player. So I want to make him the captain of the team.  
Why is this example?  
Why is this an example of a bias?  
Why do you think this is a bias?  
Why do you think the choice of a captain may be skewed?  
Yes, looking at this information.  
Why? And I want you to put on your thinking caps like these data analysts, because this is these are the things that you are going to be doing in your workplace when the data comes.  
You are not only observing for data data, you are looking for data that if it is skewed or not. If there are biases and these are some of the scenarios that are going to let you know.  
Some of these things. So Jeremiah.  
I saw your hand up. Why?  
Why do you think this is an example of a bias?

 **Jeremiah Elorm Aleawobu** 11:27  
OK.  
So what's a mistake?  
Is it was a mistake?

 **Emmanuel Kofi Essel** 11:30  
Right, yes.  
Anyone think about this carefully?  
You are only considering performances where it did not ring.  
Yes. Why do you think this is a bias?  
Why do you think this is a bias, Madea?

 **Mardiya Ananu Weyire** 11:56  
Because it's a bad because they are only assuming that the person that you are basing your decision on live very well when the sun was out.  
What if my talent comes out when the the rain is pouring?  
What if my my skills are?  
You know, when the rain is falling. I'm feeling the blessings from God and I'm playing crazy.  
So you can't judge me, so only one.  
Set of life stuff. You have to check from all angles.  
So if you judge me, then what you want then that's a bias.

 **Emmanuel Kofi Essel** 12:29  
Right, great.  
That was that was great.  
Let's let's give it up for my dear.  
So you you actually don't have to just because and most of you that is how we make decisions.  
This what we are talking about is not only even about data analysis, data analysis and all that. Most of you make your decisions based on this.  
Oh, I know him.  
Oh, I know her.  
Oh, we go way back and so.  
The the person you knew 10 years ago.  
And all that, whatever you think that 10 years is not enough for this person to change.  
Oh, I know him.  
Oh, we were friends way back there.  
Way some people are laughing. I'm not talking about boy gay relationship.  
I'm talking about you are an hour and you know the person has come for employment and you think you know him so well.  
Hey so you want to consider him? That is what I'm talking about, too.  
Don't put me in any trouble. This morning I came in peace.  
Yes. So you make your decisions based on history.  
What has happened in the past?  
Some of you.  
Because of what has happened in the past, you will not give them a chance. Think that.  
Oh wait there.  
You use your popular popularity.  
Maybe I'm not so kind and training man 'cause. I'm not so.  
So you, you, you you use history to take decisions and take decisions and before you know it, pasta that is an example of a bias. I'm telling you this morning.  
Telling you it's an example of a bias, OK.  
So let's look at this scenario.  
Rosemary is going to read for us because she's very smiley this morning.  
I'm learning the names of hairstyles so this one you have to tell me the name of.  
But you read it before, so read what is in the green first. Then you go to what is the the white page. So let's go.

 **Rosemary Akosua Akarimanga** 14:35  
So a company decides to streamline its recruitment process, season AI to screen job applications.

 **Emmanuel Kofi Essel** 14:44  
OK, then let's go to the right.

 **Rosemary Akosua Akarimanga** 14:47  
OK.  
So the air is trained on historical hiring data to identify idle candidates.  
However, the data reflects years of biased hiring practices where most employees were men. The algorithm learns from this pattern and starts favoring new candidates, rejecting equally or more qualified women and candidates of other agendas.  
This case illustrates how bias in the data or algorithm can perpetuate.  
It's inequality.  
Even when the goal is to improve decision making.

 **Emmanuel Kofi Essel** 15:23  
Awesome. So in a company where you are trying to the the goal of the whole process is that you're going to get good candidates, right?  
But the data that exists which is going to help you train your model and all that is favoring men already.  
So from the word go.  
Men are favored over women.  
And we are going to use some data to to do that.  
We are going to test it to see how it works so.  
If you follow this algorithm based on the algorithm and the data history, we are going to learn that this, even though it is a good idea to to improve decision making your data that is being used to train the model is wrong and your algorithm also is not.  
Written to identify some of these errors and so it will also go ahead and do whatever it has to do and at the end of the day, more meals will be.  
Given.  
Employment over females.  
This is bias and this one. It is not something you are going to look by just observing the bite, because by the time this data comes, it come clean before you give any data to an AI to train it, the data must be clean enough.  
And if you look at this scenario, there is bias.  
We are mentioning all this because we want you to know that as much as it's it's it's going to be a good thing to clean day-to-day time duplicates and all that.  
One of the things that you must look out for is biases.  
I saw Justina's hand up Justina.

 **Justina Arthur** 17:13  
This was a mistakeholder.

 **Emmanuel Kofi Essel** 17:18  
I hear so looking at all that you have sent up until this point, what do you think?  
Why do you think it is important to identify bias in our data?  
Or even in our process.  
Yes.  
I think I should start calling calling people.  
So let's start from Sandra.  
Sandra, why do you think?  
Identifying bias.

 **Sandra Koranteng** 17:49  
Hi.  
In order to make good decisions.

 **Emmanuel Kofi Essel** 17:54  
In order to make good decisions, I see that's one. OK Jeremiah.

 **Jeremiah Elorm Aleawobu** 18:09  
So that's the we we don't skew our results based on consideration of factor.  
But then we look at all alternatives and make a right judgment out of it.

 **Emmanuel Kofi Essel** 18:22  
Great. Thank you very much. OK, wermina.

 **Wilhemina Mensah-Doe** 18:28  
In order to.  
Accurate results.

 **Emmanuel Kofi Essel** 18:36  
In order to give accurate results, great.  
So if bias exists in your data automatically, your insights are going to be biased.  
Your decision is going to be based on your insights, which is going to be biased.  
The decision, the action that is going to be taking is going to be biased like because of a bias data, every process that is going to happen or everything that you are going to do based on that data is going to be biased and.  
That is not a good sign for a good data.  
So you must be on the lookout.  
So this morning we explain the importance of identifying and addressing bias, which we have done and we are going to be using machine learning methodology to ensure that there is fairness in decision making.  
We will use Python to explore some data sets, identify sources of bias in sensitive demographics and.  
Take steps to reduce it.  
And train and build a basic machine learning model that minimized bias and produces fair outcomes.  
So these are some of the things that we are going to do this morning.  
We have started already now.  
I want us.  
To.  
I want us to do a very short.  
Research right you are researching on some things, so I'm giving you 5 minutes to do research on these things.  
Historical bias?  
Sampling bias and algorithmic bias.  
These three types of bias I want you to do your research on it.  
I'm going to call anybody at random to explain.  
These biases to me.  
So historical sampling and algorithmic.  
Biases. Let's do that research in 5 minutes.  
It starts now 945 coming back to ask.

 **Sackitey Abdul-Letif** 21:02  
Historical bias?

 **Emmanuel Kofi Essel** 21:05  
Historical sampling and algorithm or algorithmic bias.  
Thank you.  
You have some 3 minutes.  
We have a minute more.  
OK, so time is up I want.  
One lady to tell us what historical data is based on what she discovered and then from there we want a guy to tell us about.  
Sampling data.  
So let's start any volunteer.  
If not, I'll call somebody.  
Any volunteer?  
Well, since there is no volunteer, then Adam is going to tell us.  
What?  
Adam, you are speaking, but we can't hear you.  
I see.  
Martha is going to tell us, Martha.

 **Martha Afful** 27:34  
Yeah, please.  
Which one am I speaking?

 **Emmanuel Kofi Essel** 27:37  
Historical.

 **Martha Afful** 27:40  
OK.  
So what I discovered with historical is that most of the time they're bias or caused because they've tuned the model to depend on like.  
Past issues, so maybe if there's any current issue that you try to test the model, it will not.  
It will favor the information from the past like.  
The example you gave previously about the.  
And males being favored because the mother has been trained with fantasy.  
No candidate. So even if you were a female and then you tried to apply for that job, they will just not accept you because the model knows that from our past experience, most of our employees are male.

 **Emmanuel Kofi Essel** 28:32  
OK, great.  
So based on past experience and that is the key theme.  
Past experiences history from data from history. From this could it could be that as much as the data from history after that time that data was accurate and then taking decisions based on that data was relevant.  
Today, it may not be relevant, so if you are looking.  
At only history data from history, you are likely to perform an error.  
And you are going to use.  
Some machine learning.  
To do that, now let's go to sampling bias.  
Sampling bias. A guy Bennett.

 **Bernard Doyi** 29:24  
All right, so from what I read about something by as it has to do with the fact that the technique used in sampling, OK.  
Is that such a way that it favors a portion of the population, or yeah, so, so that can be done in terms of data analysis when data is about to be collected, the sampling can be done such that at the end of the day the output or the.  
Inside out to be generated will be section section of the society.  
Leaving the other section of the society in a way that disfavors it.

 **Emmanuel Kofi Essel** 30:01  
OK.  
Great. Thank you very much.  
So you you are.  
You are in a class in a school. You want to know learners performance and.  
You are.  
You go to the class and whilst you are sampling, you sample the best students like 10 of them.  
All the learners you are taking are grade students.  
And you take 10 of them like you, you need 12 students as your data.  
You take 10 of them.  
And these stand out from your top 10.  
In fact, they are.  
They are the top 10 and then you take extra two from the bottom 2.  
Do you think this is going to represent the exact data the given the goal?  
Do you think this is going to do justice to the goal?  
Yes, give me a terms down if it's not going to, it's not going to do justice or if you think it's going to do justice to give me a thumbs up.  
You want to know Lennys performance.  
How leness are performing in a class of about 40 students. You sample 12 and in the 12 you sample there and the 1st 10.  
Are good students very good student and in fact they are the 1st 10 students in the class and you take the bottom 2.  
Do you think you're going to do justice to the goal? Thumbs up.  
If yes, thumbs up if no.  
Why you are not on my name you.  
Well, since you are not my name, me Belinda, you are going to tell me is it is it?  
A good sample to take.

 **Mardiya Ananu Weyire** 32:02  
Teacher coffee. The answers are in the chat.

 **Belinda Ntow** 32:07  
No, no, please.

 **Emmanuel Kofi Essel** 32:07  
Well, unfortunately, unfortunately my chat takes awhile to to load, so that's why I prefer the time SAP and time's down.  
But anyway.  
The answer is no.  
Now let me take this thing to.  
If you are you, you want to talk about female reproductive health, right?  
You want to talk about female reproductive health in a class. There are 60.  
Lennys.  
And there are 40 ladies, 20 guys.  
If you want to take a sample.  
And you take 10 boys and 10 girls.  
Have you taken the right sample for the for for your study for your goal?  
Yes or no.  
There are 60 learners in the class.  
40 girls, 20 boys you are looking.  
At you are looking at talking about female reproductive health and you take a sample 10 boys, 10 girls.  
Have you done justice to the goal?  
Will you achieve your goal?  
Thumbs up if a yes.  
Thumbs down. If you say no.

 **Gabriel Agbefu** 33:32  
No.

 **Emmanuel Kofi Essel** 33:36  
You can use the.  
You can use the emoji.  
Emmanuel, you have a question?  
Gifa, I'm coming, Emmanuel, who asked first. Then you coming.

 **Emmanuel Klutsey Ofori** 33:57  
Please with regards to a large, maybe population or something. So how do you go about that one?

 **Emmanuel Kofi Essel** 34:07  
As in a large population based on which of the biases are you talking about sampling?

 **Emmanuel Klutsey Ofori** 34:12  
Sampling, yeah, sampling.

 **Emmanuel Kofi Essel** 34:15  
Yeah, I think there's a mathematical formula to be able to get the correct sample. There are even.  
Online stuff that you determine, you give them your the sample size.  
No, you give them your population, you give them the metrics and it will tell you the sample size.  
Places you take so there are things there, but we will talk about it when we get there.  
So don't worry, OK?  
Yeah, that is what is, what is going to happen.  
We will talk about it when we get there.  
Jiffy.

 **Dzifa Ella** 34:54  
Please, it was a mistake.  
Sorry about that.

 **Emmanuel Kofi Essel** 34:56  
It was a mistake I see, OK.  
So based on the question I asked earlier, it will be an error for you to take ten boys, 10 girls because female reproductive health doesn't have nothing to do with the boys.  
So if you want to take a sample data and your sample size is 20 which is half of the girls, that is fine.  
At least it means that your 50% there is a 50% chance that.  
Your data is accurate. If you get a larger sample size like 30 then you are looking at 80% chance.  
So these are some of the metrics around biases and all that.  
So please take this into consideration now.  
The last one is bias based on algorithm Rene.

 **Rene Missah** 35:55  
Taco Fi, I don't know, but my listen is running, so I've not been able to get anything. But I'm here, so just go and come.

 **Emmanuel Kofi Essel** 36:02  
Your your destiny is running Sir.  
Which which of the destinies are we talking about?

 **Rene Missah** 36:09  
My my laptop, like everything is spinning and even my chat is not loading.

 **Emmanuel Kofi Essel** 36:18  
OK, I hear.  
OK.  
So we are going to.  
OK.  
Who else is going to Justine algorithm?  
Bias based on algorithms.  
I think Justine has frozen.  
Pearl, Claudia.

 **Pearl Claudia Abedi** 36:49  
Hello.

 **Emmanuel Kofi Essel** 36:51  
Yes, ma.

 **Pearl Claudia Abedi** 36:53  
So the algorithm bias is refers to unfair outcome generated by AI systems.  
Due to inherit biases in the data.

 **Emmanuel Kofi Essel** 37:09  
OK.  
So this goes back to bias based on data you gave me the wrong data, so I gave you the wrong outcome.  
And that is the bias that you are talking about.  
So if your data is wrong and you fit it into your machine learning.  
Machine learning algorithm, whatever it is that you have, you are going to end up getting the wrong output.  
So please make sure that is done. I want to.  
To test something.  
Sakiti, your hand is up.  
Sakiti.  
OK.  
So.  
Based on what we have done so far, we are going to use a particular machine learning.  
I'm going to copy the codes and put them in the chat for you.  
For those of us who can, you can follow along and do it.  
Right. But it is not.  
Something I'm going to ask answer questions on like I was setting my listing up, it didn't work.  
I was doing this.  
No, I'm not going to take those questions.  
So you have the liberty to.  
Follow along whilst I share my screen and then we do it.  
Or you can test it, so I would rather you do this.  
Let us all do it together on my screen.  
I've shared the code in the chat is going to take a while for it to come back.  
I know it's going to come anyways.  
Oh, why is this?  
Trying to paste it so that.  
Look now before you can run this, you need to import pandas and sign for sign kits Len.  
So please make sure you are doing that if you want to run this on your own.  
But.  
I I'm sharing my screen.  
There is a code here that I'm going to take all of us through.  
Please confirm. You can see my screen I'm using.  
I'm showing you Jupiter notebook, so if you can see my screen.  
Please confirm by giving me a thumbs up.  
OK, great.  
So.  
Like I said, the code is is.  
The code is in the chat.  
You can go ahead and try it on your own, but first we are importing our pandas library.  
So here we have it.  
Imports I am taking you through what each line of code is doing.  
So import pandas are speedy, which is important the pandas library.  
And before you do.  
You do this.  
You should have pandas installed already.  
This you by now you should have it installed because of course you have done.  
You have already done that.  
Through some other exercises, so I'm sure you already have.  
The pandas, the next one you have to do is to install.  
The.  
The Science kits library.  
Give me a SEC.  
Give me a SEC.  
OK, so you can install the San Kiss Library or the San Kiss Lane Library and.  
That is what you are seeing on my second line. We have done something called logistic regression.  
True or false?  
How many of us remember logistic regression?  
We have not done logistic regression.  
You have it. Oh, OK, great.  
That there are types of regressions and these types are good is we are using statistical analysis for logistic. When you use that logistic regression, usually your result is supposed to be either true or false or yes or no.  
We use this regression to determine.  
A yes or no value based on variables you have.  
Based on giving variables and that is what we are going to do today.  
So we explained that when we are going as we are going so from the library that we imported, which is the San Kitts Lane, Renee, your hand is up.

 **Rene Missah** 42:58  
Yes, difficulty. I I can't see. You said you sensed the code and this thing. I can't see anything in on the WhatsApp page too.  
I can't see anything.

 **Emmanuel Kofi Essel** 43:09  
It is in the chat and I know it's going to take a while for it to come and I intentionally did it that way so that whilst we are doing.  
Somebody else will not be trying to run the code and before we know it you you didn't get anything.  
So by the time I finish taking you through this, I think that it will show up.  
But I've pasted it in the chat, so don't worry you get it.  
OK.  
Yeah, I see it's loading here.  
People's chart from 90 clock announce showing in my chart. So you can imagine what's going on.  
So we are importing.  
Train under score test dot splits under score splits from.  
This signature model, this S key learn model.  
Is short for sign kits Len to SK length signature length.  
And there is model selection which is selecting this model.  
Then another model.  
Called linear model, also from the same case, let it's being imported and from there we are importing logistic regression.  
You understand what it does and from metrics which is also in the San Ketelane we are doing classification reports and that is what we are important.  
So we are actually importing train test plates, logistic regression and classification reports.  
Based on we are importing them from the San Ketel.  
OK, now.  
We are creating a data set.  
So this is a dictionary in Python Mata.

 **Martha Afful** 45:01  
Yeah, please.  
Can you explain what the 1st and the last one does?  
Because you only spend this account.

 **Emmanuel Kofi Essel** 45:11  
These are models you are going to use when we get there. I will explain it for you to see.  
So don't worry.  
So here we are creating our datasets called data we have.  
We are using a dictionary.  
We are using a dictionary here, so we have the key as gender, so we are taking three by when we are done, our data is going to have 4 columns, the gender, the experience, the test score and then hide.  
This is a data.  
This is the historical data database.  
On what has happened in the future.  
So when you see one, it means you are the person was hired.  
When you see zero, it means the person wasn't high.  
So this is the data we created and we are using pandas to create a data frame for it.  
Now on this, so we are saying DF pandas data frame we are turning the data that we created here into a data frame with pandas. That is what we see here.  
And for here this line we are changing the gender values you see when you come here. Gender is in strings, male, female and all that. But in machine learning the machines can only understand numbers once and zeros.  
So what we do is that.  
We are using the Lambda function to change.  
Mail the gender meal to one and female to 0 so that is the formula that has been written here. Again, the code is available when I give it to. You can use ChatGPT to run and understand these things.  
So the code is available then after all that we print the data frame to look see how it is looking like.  
So this is the score.  
You see that now gender is no more male and female.  
It is one.  
What represents one for meal, one for zero for female.  
Now look at. Let's look at the Hyde versus gender.  
Observe this and tell me what you see.  
Look at the height, the gender and then the height.  
Renee.

 **Rene Missah** 47:55  
They they are the same.

 **Emmanuel Kofi Essel** 47:59  
They are the same.  
So what is the observation?

 **Rene Missah** 48:04  
So it means.  
Gender, as in one mil.  
The hide is also one.

 **Emmanuel Kofi Essel** 48:17  
OK, you have you have seen the thing.  
Like I said, BP whenever any.  
Po when I mean that is what is happening good.  
Let me see.  
Let me get somebody who has not talked, spoken to the Mildred.

 **Mildred Tseh** 48:36  
OK, so I can see that out of the 10 people they hired there, there are fours.  
There, which means they have 4 figures who are hired and the men are about 6.  
So more men were hired than the females.

 **Emmanuel Kofi Essel** 48:54  
Females were hired.  
OK.  
That's a good observation, but I think you have also seen it in the spirit. I can say VP, when I'm in, I think I saw my dear Sandra.

 **Mardiya Ananu Weyire** 49:14  
Yes, so the output is only meals were hired. There were no females that were hired.  
That's why the gender when is 0, which is for female, the hide is 0, which means not hired.

 **Emmanuel Kofi Essel** 49:28  
Great. Thank you very much.  
So if you look at the data, well, your observation is that we see everywhere there is a male gender, their hired is true like one.  
So it means that only males were high. Look at this.  
This is a female and a female wasn't hired.  
But now look at the test call.  
Let's look at the test call.  
It is the test call that is going to make us understand how bias this thing is looking.  
Like so, I want another observation based on the test score, gender test score and height.  
Gender, Tesco and height.  
Look at it carefully.  
That is where we are going to understand the bias in itself.  
I want somebody to give me an example of a situation in this particular data that makes it biased. Bernard.

 **Bernard Doyi** 50:31  
OK, so looking at the first meal, OK, and we had 85 and he was hired, OK, now let us look at the first female. She had 88, meaning that she did better than the first male.

 **Emmanuel Kofi Essel** 50:36  
Hmm.  
Good, good.

 **Bernard Doyi** 50:45  
However, she was not hired OK.  
So this is one of the examples we can look at.

 **Emmanuel Kofi Essel** 50:52  
Exactly. Let's clap for him.  
Let's clap for him.  
That's that's a very good catch.  
Look at the when you even look, go ahead and look at the experiences.  
The the mill guy.  
The first mill guy.  
The experience is five and got hired, but a few mills experience is 6 and wasn't high.  
The the The, the one that is even shocking is the fact that the mill, the Ted.  
Mill had an experience.  
Of three and was still hide.  
But the female has an experience of six and wasn't hired.  
So you can just imagine how biased this data is.  
But this is a carefully cleaned data.  
So this is to tell you that the data cleaning and the fact that you have good and accurate data, it's not only because you have cleaned the data, it's because you are monitoring to observe biases.  
And this is what we see here.  
So now let's let me come here.  
So now we are calling on.  
We want to.  
We want to train our model.  
Abigail, do you have a question or it was a delayed hand raising?  
Any question?  
OK.  
So when you look at this, you can see that we are trying to create another from our data frame.  
We are splitting.  
Jen this ones into different targets, right?  
And then on the X axis, we are looking at gender experience and test score. And then on the Y axis, we are looking at the hide.  
So we want to compare.  
How the gender, the experience and the test score relates to being hired so.  
We are using the train model that we imported the X to train.  
And test. So we have this particular data frame we are using.  
We are going to train it and after training it we will test it.  
So we train.  
We are splitting the data into two one data set for training, the other for testing.  
So you see that X has a trained model and has a test and the same thing happens for Y.  
And test and we are using the train test plates that we imported matter. So this train test plate is used to split data the same data into a testing data set and then a training data set.  
Right. So the data you create this particular model is going to divide them into one set to be trained on and the other set to test the training.  
Data to see if it is accurate.  
I mean, I mean you seen that test size of 0.3?  
Then we are looking at which model we are going to use and we are using the logistic model, which means that we are expecting either A1 or zero, true or false, yes or no.  
We are looking at the chances that a female is going to be hired or a male is going to be hired and we are using the logistic regression model.  
Now we are fixing the data.  
We will train the X model.  
On the X we will train it and we use the Y.  
We are training both X&Y, so we are fixing it into the model.  
We are using the logistic regression.  
Again, when you go to, when you use ChatGPT, you are going to understand it better.  
But you are going to train the model.  
This one we are just fixing our data into the machine learning model.  
So that it can perform the analysis.  
Then this is the wide prediction.  
This is where we are going to predict.  
Based on the data.  
I remember the.  
X the XY and the X text here is.  
Comparing this is on the X so we are comparing the gender, experience and Tesco.  
We have already splitted another this same data for training. So after the training is done we are going to use the same data to test to see if we are going to get the accurate reports now.  
This is.  
The results?  
The chances that a female is going to be.  
A female is going to be high is .50.  
The chances that the meal is going to be hide is one.  
Which is like 100%.  
As for these ones, you may not.  
Really get an understanding, but this is where we use this data to discuss.  
So the precision that women are going to be hid is less than the precision that men are going to be hired.  
So if you look at this data, you see that based on the history of the history that we have.  
More men are going to be high over women.  
So if you want this to be done.  
Then you have to go all the way back to your data and make the adjustments.  
Please any questions on this.  
Any questions on this?  
Anyone with a question?  
Are we good?  
If you are good, give me a thumbs up.  
More more of this is going to come.  
So you understand, but like you said, pipes for Python. You can take these codes that are shared. I'm sure by now you should be showing if it is not then I can gladly share it in the WhatsApp US Now I know.  
You you get it.  
So the last thing we are going to do, we are supposed to do this in breakout rooms, but I want you to try it on your own.  
So I'm sharing.  
I'm sharing.  
A document with you what you have to do.  
Let me share with you placing it in a WhatsApp chat, so you may want to go there and check it out.  
OK.  
So that is the 1st.  
Documents it says a handout.  
And then the second document.  
That's that is a CSV file.  
Please, when you go you downloaded the instructions are in.  
The handouts that I shared.  
It's supposed to be a group thing, but I want you to do it individually so that the understanding will come clearer.  
So this is a CSV file. Please download it.  
I've seen some of these questions. Bias is always completely wrong.  
I think this depends on what you are doing.  
But I'm interested.  
Bright, what do you what do you think?  
Or when is a bias not wrong?  
Which example?  
Which situation do you think the bias is not wrong?  
I want to learn, so tell me.  
Give an example of the top of your head.

 **Bright Hodogbe** 1:00:00  
OK.  
So yeah, yes.  
So coming from my background, I know that sometimes when student performances rights these lectures, when the market scores end, they expect it to always be a normal fellowship care distribution rights.  
But then you realize that maybe you see that some of the majority of the students will probably be way lower than what you expect.  
Or they carefully skewed so, and then you find maybe one or two outliers.  
Somewhere, right?  
So what they end up doing is to add app scores or maybe try to massage the data so that it can like come back to the middle and probably have that dumbbell shape.  
That is some accent, right? In such a situation, we know that the original thing is already skewed, which is bias, right?  
But then there's the case in order to.  
Let's say let it come to the standard of the university.  
But maybe that's the kind of Lord or rule that is that the results should always have like that kind of \*\*\*\*\*\*\*\*, stuff like that.  
So in certain situation we consider the massager is that 'cause. That's also another form of bias, right?  
Because you're also taking the data and to shift from the skill to the normal shape.  
Would that be considered?  
That's what? That's what I'm asking.

 **Emmanuel Kofi Essel** 1:01:23  
Why are you not?  
Why don't you decide to introduce more data instead of?  
More variables instead of skewing the already existing.  
Because if you skew it or change the data you are not representing the right the accurate one.

 **Bright Hodogbe** 1:01:43  
Yes, that that's the point.  
I'm saying yes, that's the point.  
But if they do that, what ends up happening is like the performance of the students would be so abysmal that they it's not something they encourage to be recorded.

 **Emmanuel Kofi Essel** 1:01:59  
I see.  
I get what you are saying, but honestly for the for the benefits of the lecture key is better off.  
Maybe he can present the original 1 and skew the other one to fit whatever purpose he wants to do, but that is not accurate.  
And if you take a decision based on the skewed one, the one that you apply some biases and added some data, it means that you are not going to.  
Have consistent.  
Results right?  
Your your you may get a good result, a solid result based on your data analysis, but the results are not accurate because originally the lengths were formed abysmally in your own words, but and so when you are taking a decision, your decision must be based on the AB.

 **Bright Hodogbe** 1:02:44  
Thank.

 **Emmanuel Kofi Essel** 1:02:48  
Performances and the fact that that you develop strategies to help learners.  
Based on that.  
But when you skew it so that you can go and Get the facts from managers.  
Manage management meetings and all that and they take a decision based on that. They have not held the whole school.  
Right. So it it depends on the situation, whatever it is that you want to do. If it is not a decision that you just want to have a fun of it. Or maybe.

 **Bright Hodogbe** 1:03:11  
OK, I get you.

 **Emmanuel Kofi Essel** 1:03:25  
You are.  
I don't know lie if I should say then well, you add your bias, but you, you and I know that the original data, the correct information is that the Linux are not doing well.  
So if you have to do something about it.  
You do have about it.

 **Bright Hodogbe** 1:03:42  
OK.  
My next question is this, so I remember you mentioning about the selection of people in a class for, let's say, reproductive health.  
Something like that.  
I mean, I didn't get to, but then I was thinking right.  
I know that when it comes to sampling, randomness is encouraged, right? Random selection, right?  
So if I come to a class and I need, let's say, just some people to do a study, right?  
I could say, oh, OK, can I?  
How many of you are willing to?  
So maybe partake of this survey right is on this right?  
And probably there are some basic questions. I would like you to answer.  
Raise your hands. Or can I just select any of you and then you partake of it?  
I mean, that's randomness, right?  
And how? I mean if I select and then I end up having maybe more guys, right?  
And it's a lady state, I mean.  
Will it be my fault for generating that introducing that bias in the?

 **Emmanuel Kofi Essel** 1:04:46  
Yes, yes and yes. That, yes, this time is a capital Y capital E capital S with three exclamation Marks and maybe maybe 10 full stops. The thing is whatever you are going to do based on your goal, the randomness should come from the ladies. You have 40 you.

 **Bright Hodogbe** 1:04:47  
Thank you.

 **Emmanuel Kofi Essel** 1:05:06  
Are taking 20, so 20 random ladies.  
Can raise their hands.  
Reproductive female reproductive health.  
What? What data are you taking from the guys do you like?  
What questions will you ask the guys?  
Do you?  
Does your breast grow?

 **Bright Hodogbe** 1:05:24  
Or maybe.

 **Emmanuel Kofi Essel** 1:05:27  
Does your hips go down?

 **Bright Hodogbe** 1:05:29  
So. So for for instance, assuming it's on a topic like let's say breast cancer or anything or breast, right?  
Maybe you can ask the guys.  
Maybe they are people who are married.  
Have they an incentive to have they? Maybe.

 **Emmanuel Kofi Essel** 1:05:42  
Class 3 Moffat All class ADNS your friend, say JSS 104 who is married.  
Well, I I know.  
I know that you want to establish the fact that the the the randomness part, but I'm just giving you the clue that yes, randomness is true.  
You need to be as random as possible.  
It shouldn't be an intentional sampling, right?  
Just so that you get the the fair idea of what is happening.  
But you already have a target group.  
You are talking about female reproductive health.  
Female reproductive health.  
They they did tell you must have must be feminine like it should be women responding.  
It is only like when we are talking about topics like football.  
And we select only men. That one, it is not fair because we know some women like football too, so that will not be a fair thing.  
But we are talking about female female reproductive health and go and what is the guy coming to do?  
Like what question will you want to ask the guy?  
So that you get your data because already before you even go you have the set of questions you are going to ask, right?  
When was your first menstruation?  
How? What was it?  
They all those things and what's answer? Will a guy give you?  
When was the first question?  
So if you take 10 boys and 10 girls, it means that automatically you are going to get 0 for that question for all the boys.  
Which could have been.  
Different.  
If they were girls.  
Right. And that will represent a more accurate.  
Peter done taking the boys because he want to be as random as possible. If it is a particular like you are too specific on your goal, don't waste your time going to check somebody because like your adding somebody because you want to be as random as possible. I.  
Don't. I don't know what like I don't know what you use the men to do.  
Guys please, this is where we we end the session OK.  
Do the assignment that I have given you.  
You try your hand gets the data, you can't. You look at the data.  
Look at how you are going to use AI to help you model identify errors.  
Identify the type of error.  
Use the AI to help you identify what type of biases existing in the data and suggest have a conversation after that, copy your interaction with AI.  
Save it in a template and.  
The the.  
File after you have edited it to save it in that template and share it in the link. I'm going to share the link with you by is going to go to the full drive that we submit the assignment in. So just in case you have forgotten, I'll share.  
Could you somewhere along the line so we are supposed to go for break now?  
Yes, I think we have break from now.  
But we are supposed to have the break from 10:15 by the way. So we have already done 13 minutes out of.  
The break.  
Sorry about that.  
It's not.

 **Bright Hodogbe** 1:09:15  
Will say now.

 **Emmanuel Kofi Essel** 1:09:17  
Yeah.

 **Bright Hodogbe** 1:09:19  
We still don't have the link to the Jupiter notebook for you.

 **Emmanuel Kofi Essel** 1:09:26  
It is not Jupyter notebook. I sent the code just quotes.  
So the code is going to appear in the chat.  
We have flex time crial from now till 10:45. Then next session starts at 11:00. Oh why was I even word?  
Then we can do the this thing now.  
So all of us should do it and then submit it, Renee.

 **Rene Missah** 1:09:52  
Geography. I was saying that you said you we should run the code but the code never came.  
So maybe if you can put this on the WhatsApp.  
Or that's what you are about to talk about.

 **Emmanuel Kofi Essel** 1:10:04  
Yes, yes.  
But don't worry.  
Let me put it in.  
The this thing for you?  
It is only the.  
It is only the.  
This is the first part of the code.  
And.  
This is the second part of the.  
So let's try our hands on it.  
Not only could do on.  
The group work that we're supposed to do, that I want you to do it individually. Let's try.  
Hands on it.  
Open the handouts. Use AI to.  
Work on the data.  
And let's get the result.

 **Mary Duodu** 1:11:25  
Hey, Jeremiah, can you hear me?

 **Emmanuel Kofi Essel** 1:11:28  
Yes, I can hear you.

 **Mary Duodu** 1:11:30  
I think that class is still recording.

 **Emmanuel Kofi Essel** 1:11:38  
Oh yeah.

 **Emmanuel Kofi Essel** stopped transcription