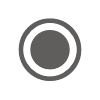
**Transcript**

December 19, 2023, 12:31PM

 **Du, J. (Jinrui)** started transcription

 **Du, J. (Jinrui)** 0:04  
So doctor Massetti in the recent five years, you work as a lead scientist in statistics at BASF.

 **Marcos Malosetti Zunin** 0:13  
Yes.

 **Du, J. (Jinrui)** 0:13  
I'm sure you have led several projects.  
You have visions.  
Goes for these projects.  
Could you tell us about typical visions and typical subgoals in order to achieve the long term vision?

 **Marcos Malosetti Zunin** 0:31  
Let me see your question is.  
As leader of of certain activities that connects with projects, what is the vision on on those is uh.  
So yeah.

 **Du, J. (Jinrui)** 0:48  
You have, yeah.  
You you had.  
You are the lead scientist and you have that several projects.  
And I'm sure you have the each project has its vision and each project has goals.

 **Marcos Malosetti Zunin** 1:05  
Yeah.  
Umm.

 **Du, J. (Jinrui)** 1:13  
So what are the?  
What is the typical vision for your project?

 **Marcos Malosetti Zunin** 1:17  
Right.  
OK.  
Yeah.  
Yeah.  
So maybe there the the the application is OK, they have we we work with project yes, but actually and we work more in, in, in in the frame of what are the the the company goals or the the function goals yeah.  
And and then every year this these goals are discussed and planned and according to that then we plan our activities.  
Yeah, it's A and.  
And of course there would be project involved on that.  
But say the way to to work and plan our work is in alignment with the whole organization.  
Yeah.  
And of course, we are asked about what is our vision.  
For example, when it comes to the to the to the area on on how we see data analysis, yeah, which are the areas where we should explore because we see opportunities.  
So how we deploy our data analytics internally that that we have a vision and how we think that should be done?  
Uh.  
And then there are larger projects or activities or something programs you know collection of different projects that that have a shared vision, not only our own but actually shared vision.  
And then we contribute to that, yeah.  
And.

 **Du, J. (Jinrui)** 2:40  
Yes, yes.  
So the visions are not from you, but uh, the whole company discussed this.

 **Marcos Malosetti Zunin** 2:47  
Exactly.  
There should be an alignment and then of course we have our our visions when it comes to to our domain, the knowledge domain area, yeah.

 **Du, J. (Jinrui)** 2:56  
So in your dog you are knowledge domain.

 **Marcos Malosetti Zunin** 2:56  
But the vision, yeah.

 **Du, J. (Jinrui)** 2:59  
Can you tell me, tell us about your goals for, I mean, yeah.

 **Marcos Malosetti Zunin** 3:03  
Yeah.  
Well, exactly.  
We we we as a group, we are a global group.  
So we are a group that actually is located in different regions from from the Americas to Europe and and and Asia and and what scope we have is OK, we have to deploy data analytics to every single stakeholder in the company wherever they are and and and.  
Uh, we know we cannot do it at manually we we we should we should automate this and how we make that in an efficient way and that's what our regional camps and say how we make possible that we put state of the art technology in terms of data analytics without having asked to intervene directly in those data analysis.  
And that's that's, say, one of the of the component and the other very important component is how you communicate that to the users because it's not about the technology by itself.

 **Du, J. (Jinrui)** 3:55  
That sounds, yeah.

 **Marcos Malosetti Zunin** 4:09  
It's about how people can use it and and to be use it possible to be use it.  
Then people should understand it and we doubt having to know every single detail of of every technical aspect in the in the statistics to lead, to give you an example.  
I mean, we are working with mixed models all the time, so that's our core technology.  
Uh, but a breezer or someone at this in the decision making doesn't need to know and we do not know what Dremel is and and actually we should not be explaining what REL is, but actually how to interpret these results to guide the decisions.  
And why does it?  
Does it make a difference?  
Uh, so that that's also part of our vision is OK how to develop state of the army theology because he also how to make it accessible to people and understand why this is important.

 **Du, J. (Jinrui)** 5:10  
That sounds amazing.

 **Marcos Malosetti Zunin** 5:11  
Hmm.

 **Du, J. (Jinrui)** 5:12  
Thank you for your answer.  
And the next question is about life.  
As a researcher, I will pass you over to my teammate Sonja.

 **Marcos Malosetti Zunin** 5:22  
OK.

 **Song, J. (Jia)** 5:23  
Yeah.  
So sorry for the late because I have problem with my camera.  
I don't know.

 **Marcos Malosetti Zunin** 5:28  
No worries.

 **Song, J. (Jia)** 5:29  
Do you hear that I mentioned before?  
So the the second question is what do you do on your team do in each phase to achieve the goals and what does a typical day look like for researcher and how do you divide your tasks?

 **Marcos Malosetti Zunin** 5:46  
Yes. Yes.  
So uh, that's that's an interesting question because I mean I I mean in some other meetings also with, with, with students.  
And then I was asking myself the question, how do you imagine a day for us?

 **Song, J. (Jia)** 6:03  
Umm.

 **Marcos Malosetti Zunin** 6:04  
So how much do you think you we should be sitting behind the screen doing coding and doing analysis?  
Uh, and it could, yeah.

 **Du, J. (Jinrui)** 6:12  
For me, I guess you will have a lot of meetings and you called, they're little.

 **Marcos Malosetti Zunin** 6:21  
So yes, exactly.  
I mean a big part of our time actually comes with communication part.  
So talking with people, interacting with people and and and explaining what the how to use the the data analytics, uh and then I could also part is part of our time had to go into how to umm uh deploy that more than actually doing this analysis is how to make this analysis done in an efficient way.  
Uh, and part of us.  
Our time should go into scouting for what is there outside that we are not using, so they're on one hand we have something that is in production.  
So then it's being used by end users, but in parallel.  
Actually we have to develop and scout for.  
Potential new improvements that we might want to bring in uh, that also should occupy certain amount of of of of our time.  
So it it's a big part about talking with people.  
Wait, you and you said, but also internally on on, on, on, on decision making and where where which direction to take and and but but not?  
But also sculpting for innovation, that's also something that we have to, uh, to, yeah, to spend time.  
And of course, within the team there are different balances.  
I mean my case I have also work to do in terms of of management people.  
Uh, because that's just that's so important.  
I mean, we should take care.  
Actually, people stays in the job motivated and and growing.  
Yeah.  
So then, uh, that's something that actually is a manager I have to also take care that everyone finds it bright place in the in the in, in in the job.

 **Song, J. (Jia)** 8:22  
OK.  
Thank you.

 **Du, J. (Jinrui)** 8:29  
Yes.  
So you mentioned about the scouting, deploying and management and I I assume that you.

 **Marcos Malosetti Zunin** 8:35  
Yeah.

 **Song, J. (Jia)** 8:35  
Right.  
Honey.

 **Du, J. (Jinrui)** 8:39  
I assume that you there is also teams for data.  
So how do you prepare the data?  
How?  
How are the data collected?  
How much data do you work with?

 **Marcos Malosetti Zunin** 8:53  
Yeah, that's, you know, the in data analysis, the matching numbers are how much time do you spend in, in, in data quality control and a data wrangling and those things.

 **Du, J. (Jinrui)** 8:53  
Here. Yeah.

 **Marcos Malosetti Zunin** 9:07  
And normally people say 80% of the time spent on that and 20% in in data analysis.

 **Du, J. (Jinrui)** 9:10  
Yeah.

 **Marcos Malosetti Zunin** 9:13  
Uh, yeah.  
Well, that, that's that's kind of the challenge always.  
How to minimize that part of the time data is collected in various locations, not the necessarily always with the same technology.  
Of course, we're highly digitalized.  
And in the way we we collected the data and data goes into databases.  
So we access the information through databases that that gives minimum frame, but even then we are working with with real life biological systems that brings along all kind of constraints that information is missing confusions and people are working.  
So all these things brings that the data requires some, some, some additional work before going into the proper data analysis part.

 **Du, J. (Jinrui)** 10:07  
So 880% of time is data wrangling and data come from various locations.  
How much data do you work with?  
Really, really large.

 **Marcos Malosetti Zunin** 10:18  
Well, you have to think of, but we have maybe, UM, uh, fifty or more reading programs, different being programs.  
Uh, look at in different locations.  
So we're talking about hundreds of of experiments.  
Uh, different type of experiments, different type of crops.

 **Du, J. (Jinrui)** 10:39  
Ohh.

 **Marcos Malosetti Zunin** 10:42  
So it's a highly diverse data set data that we have to deal with and that's a bit of the challenge and how to create pipeline, some data analysis workflows that are robust to all the diversity because we cannot spend a lot of time to to, in detail taking care of of every single data set.  
So we need to find ways that that, that, that this time in an efficient way.

 **Du, J. (Jinrui)** 11:09  
Pipe.  
That's yes, yes, I got it, so.  
How do you then?

 **Marcos Malosetti Zunin** 11:17  
Just just for you to to to realize, I mean in the company we have 24 different crops.

 **Du, J. (Jinrui)** 11:22  
Yes.

 **Marcos Malosetti Zunin** 11:23  
With different biologies with different type of.  
Yeah, the products for some crops like tomato is a fruit.  
Yeah, but in leak is a plant or an onions.  
Is the bulbs root or carrots?  
So this is requires different type of data analysis as well as the type of trait that are measured that we have continuous trade.  
But we have lot of scales, categorical traits, binomial of binary kind of traits.  
So that that give Bing a lot of diversity and requires a lot of ingenuity, how to deal with those different type of data?

 **Du, J. (Jinrui)** 12:07  
All those data are not collected manually, they are collected electronically using devices.

 **Marcos Malosetti Zunin** 12:14  
Yes, as much as possible. Yes.  
Yeah, but we also do a lot of experiments in remote areas and and uh at grower sites, not necessarily in research stations.

 **Du, J. (Jinrui)** 12:16  
Yes.  
Thank you for your.

 **Marcos Malosetti Zunin** 12:27  
Quite a lot of that is done on farm trials.

 **Du, J. (Jinrui)** 12:29  
Satellite.  
Do you also use satellite?

 **Marcos Malosetti Zunin** 12:35  
Satellite data.

 **Du, J. (Jinrui)** 12:36  
Yeah.  
Ah.

 **Marcos Malosetti Zunin** 12:40  
These are, you know, this is a bit different than field crops where you have extensive areas.  
Uh, a lot of crops are in in protected areas or in greenhouses, so it's a different kind of situation.

 **Du, J. (Jinrui)** 12:53  
Umm.

 **Marcos Malosetti Zunin** 12:57  
Like for example, if you talk about Mace or soybean, which are really huge, huge early extensions where you can fly drones or things like that.  
Those satellites?

 **Du, J. (Jinrui)** 13:09  
Yes, yes.  
OK, sent there.  
Once you have data or ready, it's next step is just, umm, still statistical analysis.

 **Marcos Malosetti Zunin** 13:15  
Umm.

 **Du, J. (Jinrui)** 13:19  
I will hand over to my teammate.

 **Marcos Malosetti Zunin** 13:21  
Yeah.

 **Li, X. (Xiang)** 13:23  
Yeah, I I think, uh, it work in your research.

 **Marcos Malosetti Zunin** 13:25  
Yeah.

 **Li, X. (Xiang)** 13:28  
Statistical statistical method will be an important part, so I have some questions about the statistical methods or the first one is that which part do you think is the most complex or most time consuming part of when you apply statistical methods or other data science skills?

 **Marcos Malosetti Zunin** 13:36  
Yeah.  
So is that so?  
Again, yeah, your your question is, which are the challenges in terms of statistical methods?

 **Li, X. (Xiang)** 14:02  
Yeah.  
Which which part do you think is the most complex or or most time consuming?

 **Marcos Malosetti Zunin** 14:06  
That.

 **Kanbar, M. (Mikdad)** 14:10  
Yeah.

 **Marcos Malosetti Zunin** 14:11  
Yeah.  
OK for for complex and time consuming for for complex what I would bring in is OK and we're working with genetics and what it means.  
We are typically exploring uh populations and that brings automatically the concept of random effects in our models.  
And that's the reason why mixed models is such an important part of our data analytics, yeah.  
Uh.  
At the same time, we are working with different type of traits because makes models and continue straight goes very well along.  
Then we can apply that, say, we can cooperate information on on parent test, so relationships that means required of modeling correlations and Co variation in the in the data.  
That's clearly another reason why mixed models are there.  
But we also need to account for other kind of responses which are not continuous, could be binomial.  
Try it.  
Imagine counting.  
Yes.  
How many plants have been affected by these X uh.  
Also, category categories same certain properties of plants rather than putting it on a numeric scale, you put it in in categories, uh, from from certain ordinal scale, if we like it, but also requires thinking, OK, how to analyze this type of of traits.  
Uh, and and and also another efficiency part computationally.  
Many times the results need to be in a short period of time and decision had to be made and very soon after the data has been produced, product provided.  
So data analysis had to be fast and there is where the methodology should be such that actually trace efficiency with the robustness of the of the methodology of of the of the method we need to have a good analysis for making good decisions.  
But we don't really need to go to the ultimate more elaborate more because we simply don't have the time to do that.  
But we need to find the right, the right balance because to simple models is also not a solution.

 **Li, X. (Xiang)** 16:44  
OK.  
Yeah.

 **Marcos Malosetti Zunin** 16:45  
Yeah, I always say actually I don't want to wait for a whole day for some analysis to be done.

 **Li, X. (Xiang)** 16:45  
Thank you.

 **Marcos Malosetti Zunin** 16:50  
There's far too much.

 **Li, X. (Xiang)** 16:52  
OK.  
Thank you.  
Uh and I, I have one more question.

 **Marcos Malosetti Zunin** 16:54  
Yeah.  
Yeah.

 **Li, X. (Xiang)** 16:57  
As we all know, the most common purpose of and applying statistical models in a program can be a projecting future data.  
A classifying target objects or finding effective factors and many others.  
So I want to ask what is the main purpose of statistical model in your research?  
Or in your work.

 **Marcos Malosetti Zunin** 17:26  
So the main purpose I would say, uh breathing is about prediction.  
So and prediction with small data.

 **Li, X. (Xiang)** 17:32  
6.

 **Marcos Malosetti Zunin** 17:35  
Because what's what we're doing in breeding is create a new genetics, new combinations, and with a relatively small sample.  
Uh.  
Try to find out which are the ones that in future are gonna be potential new commercial varieties.  
Uh, and and the whole process takes time.  
Yeah, so developing a new variety.  
It can take easily 8 or 10 years.  
Uh, in the in the whole process?  
But then, but the moment that you have a a new potential product, a hybrid, for example.  
Yeah, it happened already.  
A lot passed already.  
Quite a lot of time.  
Then it goes into the field and this has to be tested and find out whether it's gonna be a good material or not.  
So it's predicting in the end and we're not interested in, in, in, in, in the goodness of fit.  
So it's explaining what happened in that particular trial.  
We want to from this trial make a good assessment.  
What in future that material would do because we have to narrow down from, say, hundreds to a handful, one or two, that will eventually be a a good commercial variety and some prediction is very central and having as as good prediction as possible, it's the the main, the main challenge.

 **Li, X. (Xiang)** 19:10  
OK. OK.  
Thank you.  
Answer.  
Does anyone else have other questions about methods?

 **Marcos Malosetti Zunin** 19:17  
So again.

 **Li, X. (Xiang)** 19:18  
Ohh I'm asking my teammates.

 **Marcos Malosetti Zunin** 19:20  
Huh.

 **Song, J. (Jia)** 19:21  
Yeah.  
Yeah, I'm here.  
So so Professor, can you just could you share share an instant where statistical funding positively impacts people's life?  
For example, a rail, a protection could effects on in one packs of a people or something.

 **Marcos Malosetti Zunin** 19:46  
An example of which has a positive impact.

 **Song, J. (Jia)** 19:50  
Yes.

 **Du, J. (Jinrui)** 19:53  
In fact, peoples lives.

 **Marcos Malosetti Zunin** 19:55  
Ah.  
Of our business, for example.

 **Du, J. (Jinrui)** 20:01  
Yes.

 **Song, J. (Jia)** 20:02  
Yeah.

 **Marcos Malosetti Zunin** 20:04  
Well, for example, I mean you can think of very concrete examples where we can looking at the data, provide advice on doing umm I improving the layout of for example, we're try we're discussing now of of an experiment where a large set of potentially new.  
Hybrids are tested in an efficient way, yeah, because there's always a restriction in terms of the the resources that are available and how we can make this more efficient.  
And and there is where for example our understanding of of experimental design on on on randomization, how to to allocate treatments in a a layout in efficient layout can bring to to efficiencies.  
For example, reducing this the the the whole uh experiment, so reducing the number of replications, for example, because there's not really a big difference.  
For example, yeah.  
But based on the information that's coming from, from from the data.

 **Song, J. (Jia)** 21:23  
OK.

 **Marcos Malosetti Zunin** 21:23  
So all the area of experimentally science, an area that actually brings a lot of of possibilities to to to gain efficiencies.

 **Song, J. (Jia)** 21:24  
Thank you.

 **Kanbar, M. (Mikdad)** 21:24  
Though.

 **Song, J. (Jia)** 21:35  
And OK.  
OK.  
Thank you.  
So the next question for Mikdad.

 **Kanbar, M. (Mikdad)** 21:45  
Yeah.  
So, doctor, you have a lot of years of experience in the academia and then you've made the transition to the industry.

 **Marcos Malosetti Zunin** 21:54  
Yeah.

 **Kanbar, M. (Mikdad)** 21:54  
We are pretty interested in knowing if there are any differences between those two areas.  
As a technician. Yep.

 **Marcos Malosetti Zunin** 22:02  
Yeah.  
So I mean, before before being in this in this role, I I I work in the in the and and in a group of statistics group and there and hey the type of work we did then there was very much connected with with application.  
So it was statistics, but they applied and in that sense it didn't change a lot because uh, still here working on on appliance statistics.  
And keeping.  
Keeping in touch with with the new developments, the interval methodology is very similar methodologies and applications.  
So from that perspective, no, it did not change a lot.  
Uh, OK, from the other side then it's slightly different.  
What the priorities are?  
What are the the the?  
Yeah, the the, the, the, the development.  
But yeah, I did not feel a a large different.  
Of course, I I mean well, I used to do for example lot more on on on, on the publishing, but publication for example it was used to be an editor, but now I'm not an editor anymore.  
But it's still do do reviewing, for example, because that's another way to keep in touch with with what is going on at the same time to contribute to the field.  
So yeah, with, with, with.  
Yeah, some differences.  
Overall, I would say it was not a a large change for me, but of course this is a personal situation, right?

 **Kanbar, M. (Mikdad)** 23:45  
Yeah, yeah.

 **Marcos Malosetti Zunin** 23:45  
Because if if it would be working in more theoretical area in in statistics than than I would have been different.

 **Kanbar, M. (Mikdad)** 23:52  
Yeah, I see.  
But for for example, you have argued mentioned one very important skill for for us that decision is the ability to convey ideas to other people and communicate with other people.

 **Marcos Malosetti Zunin** 24:02  
Yeah.

 **Kanbar, M. (Mikdad)** 24:05  
Yeah.  
Imagine in in the academia you might explain it to other researchers, but in the industry was it more challenging to to explain ideas to non statisticians or people that don't have the background?

 **Marcos Malosetti Zunin** 24:21  
Well, but that's that's part of of of of yes, it's it's true.  
Then a big part of of of the challenge is how to communicate this and it's different when when you're communicating that with with statistics.  
So students in statistics, of course you want to go into a lot more detail than actually when you are trying to communicate with with end users.  
Uh, so yeah, that, that, that part is different.  
Yet the communication part still remains there because, let's say as a teacher or someone working in academia, we also tend to overestimate what people know on the other side.

 **Kanbar, M. (Mikdad)** 24:52  
Umm.

 **Marcos Malosetti Zunin** 25:02  
And people are students.  
Do and the same it's it's on this side because then maybe we always need to find out.  
OK, there is the right balance and what are the key messages you want to to to convey and to make it efficient?  
Because if not, if people do not perceive this, say something that is helping, they won't simply not adopted.

 **Kanbar, M. (Mikdad)** 25:27  
So in communication was using visualization techniques and important part of conveying ideas to other people.

 **Marcos Malosetti Zunin** 25:34  
Yeah.  
Yes, yes, that that's A and.  
And that's also part of of where actually we as a statistician, we are not perfectly.  
Uh, yeah, we don't necessarily have all the skills to, for example, do developments that actually we need to work together with other people.

 **Kanbar, M. (Mikdad)** 25:50  
Yeah.

 **Marcos Malosetti Zunin** 25:55  
I mean in, in, in for example an important component on on on data analysis is communicating the results and there is where for example dashboards enters quite importantly, uh uh.  
But there we have to be aware of our own limitations and we need to work together with other people that have the right skills to put together.  
For example, very attractive and and efficient and user-friendly dashboards.  
And that's not something that statistician necessarily has.  
All the this the this this case, but it's really is a team work with with different skills.

 **Kanbar, M. (Mikdad)** 26:34  
Yeah, I see.  
Umm.  
So let's wrap up with with one final question for for me as someone who's dream is being a successful data scientist and now I'm studying Master of Statistics and data science and my colleagues are the same bot, what would be your advice for us?

 **Marcos Malosetti Zunin** 26:40  
Yep.

 **Kanbar, M. (Mikdad)** 26:54  
What's the most important skill that we need to work on in in order to be successful data scientists?

 **Marcos Malosetti Zunin** 27:01  
Uh, first of all, I mean to to to like the area.  
So enjoy the doing that work.  
Uh, the other thing is to to have enough understanding and or.  
It's not that understanding is just ask the right questions to understand the context, because you know what happens.  
But in my background I'm agronomist as a background, so that gave me a insights on the applications.  
Uh, but it's inevitable.  
People would just tag people and then if you're tagger, is the statistician, you are automatically the person that only knows about crunching numbers.  
But in in in in the.  
And the data scientists, scientists that actually is tend to communicate the importance of the statistics and applying it in the concrete application.  
Then you need to understand what the what the context is about to make it life and understanding for others.  
So, uh, my point, I I guess what the tried to make is actually it's important to connect with the users that you are interacting with and and and stand not only the that the the content of the statistics part but actually the application of course I'm talking about more from an an applied perspective.  
Yeah, that might not be always the case, but in the case of applied and it's, it's important to connect with the people that you are.

 **Kanbar, M. (Mikdad)** 28:26  
Yeah.

 **Marcos Malosetti Zunin** 28:33  
Yeah, working with so the communication is an area actually the soft skills are very important for a good statistician.

 **Kanbar, M. (Mikdad)** 28:35  
So so yeah.

 **Marcos Malosetti Zunin** 28:41  
So it's very important to have very strong understanding of the domain statistics, but also the soft skills are not to ignore.

 **Kanbar, M. (Mikdad)** 28:51  
Yeah, sure.  
So you would say passion and domain knowledge would be the most important skills, along with soft skills.

 **Marcos Malosetti Zunin** 28:57  
Yes, yes, yeah.

 **Kanbar, M. (Mikdad)** 29:00  
Yeah, this is very interesting.  
Umm, well yeah, my colleagues.  
Do you have any other questions for Doctor Marcus?

 **Du, J. (Jinrui)** 29:09  
Yes.  
Yes.  
Uh, Doctor Marcos, you mentioned many times that about the efficiency, both in terms of data collection and statistical analysis.

 **Marcos Malosetti Zunin** 29:15  
Umm.

 **Du, J. (Jinrui)** 29:19  
I can imagine that you can make a pipeline to make the data collection more efficient, but how do you make data statistical analysis more efficient?

 **Marcos Malosetti Zunin** 29:30  
Uh.  
Yeah.  
And that's, you know, any data analysis I think is in my views is is a is a decision making process where you have many different options and and sometimes or many times there's not really a wrong right.  
It's simply the option you take and what it should do is to make it transparent.  
OK, this is the option I take.  
I took and and and that is the essence of of building up these pipelines.  
It's put it together steps in the data analysis that is coherent, that makes sense.  
And this has criterias.  
For example, I was putting the criteria of efficiency, robustness.  
I don't want to have a pipeline that actually requires to fit so many parameters that actually will highly with highly likely have instabilities because that will not help me.  
I will go for a moral.  
Actually, it's robust and I'm sure that they will not break.  
Uh, as frequency as others, for example, knowing that data will not be perfect.  
So then we have to anticipate what kind of problems do data might have?  
How to diagnose this and find the decisions that actually can help fixing those on the fly?  
Then when we have 80% of the analysis in that way, I can live with 20% where actually we have to champion and data and and dig in a little bit more in detail and and troubleshoot in more detail.

 **Du, J. (Jinrui)** 31:03  
So you're saying that the 80% of the problem is solved by the by flight fixing the date data problem?

 **Marcos Malosetti Zunin** 31:16  
Yeah.  
So that that, that would be my my idea.  
And then doesn't necessarily the the the case that is always like that.

 **Du, J. (Jinrui)** 31:25  
Thank you.  
Thank you.

 **Marcos Malosetti Zunin** 31:26  
Yes.

 **Du, J. (Jinrui)** 31:26  
And you also mentioned about scouting, how do you do scouting the new new technologies?

 **Marcos Malosetti Zunin** 31:29  
Umm.

 **Du, J. (Jinrui)** 31:32  
Just reading, yeah.

 **Marcos Malosetti Zunin** 31:33  
Well, I yeah, I mentioned that I I keep try to keep in touch with with with the developments by by staying in touch with and and.  
People that actually get could be doing reviewing, for example or in conferences.

 **Du, J. (Jinrui)** 31:48  
OK.

 **Marcos Malosetti Zunin** 31:48  
Uh, that's that's one way of of of keeping in touch with developments, of course, seeing what is out there in the literature, and that means dedicated time for doing that.  
Yeah.  
And then explicit in our in our time, my location, we we we try to prioritize that in month of time to do this this kind of activities.

 **Du, J. (Jinrui)** 32:01  
Hmm.  
Thanks.

 **Marcos Malosetti Zunin** 32:13  
Right, but you need to read papers.  
We need to be in conferences at this at this kind of things that actually that's the way to to to stay in touch and also establish connections with academia and and other organizations and in in, in collaborations because it's not always makes sense to develop everything in house.

 **Du, J. (Jinrui)** 32:16  
Yeah.

 **Marcos Malosetti Zunin** 32:34  
Sometimes it makes more sense to to work together in partnerships or or collaboration projects with external partners.

 **Du, J. (Jinrui)** 32:45  
I agree with you.

 **Marcos Malosetti Zunin** 32:46  
You want to stay you in touch.

 **Du, J. (Jinrui)** 32:48  
Yes, yes, yes.  
Uh, you mentioned about 24 crops and you also say that in statistical analysis you work with small data.

 **Marcos Malosetti Zunin** 32:54  
Hmm.

 **Du, J. (Jinrui)** 32:58  
Is that a contradiction or do you?  
What do you mean by small data?

 **Marcos Malosetti Zunin** 33:02  
Well, no.  
If I smile data they I mean uh.  
The number of of experiments you can run you you can run uh.

 **Du, J. (Jinrui)** 33:09  
So N is less than P.  
Yes.

 **Marcos Malosetti Zunin** 33:16  
The per year is limited and you cannot spend 10 years trailing your material to find out this paper material because in 10 years time then maybe things changed already in terms of what is needed.  
So you have to quickly come to the market with material that can can, can can be a good material commercially, but in a in a period of time of trialing testing which is not too long in in a couple of years.

 **Du, J. (Jinrui)** 33:30  
Hmm.

 **Marcos Malosetti Zunin** 33:48  
And then there's a limited number of trials that you can run, and actually you can afford.

 **Du, J. (Jinrui)** 33:51  
Yes, yes, sweet.  
So it's efficiency is really important.

 **Marcos Malosetti Zunin** 34:00  
Yeah.

 **Du, J. (Jinrui)** 34:00  
And.  
I guess there's no more Christmas for me.  
Through the rest of the team has and more questions maybe about mixed data mixed mixed models.

 **Li, X. (Xiang)** 34:18  
I I have no more question too.

 **Du, J. (Jinrui)** 34:24  
I will be using mixed models.  
Could you give us more examples?

 **Marcos Malosetti Zunin** 34:28  
Ah, and what the reason why I mentioned mixed models? Because.  
A essentially what we're looking at is a populations right and and and there's a populations of of individuals that are highly related genetically.  
So with the typically modeling performances by modeling genetic effects that are related not it's in challenge because, UM, the material is actually connected by pedigree, so pedigree information is it's it's important in the moment of another analyzing the data and that all comes quite naturally in the in, in the mix model framework.

 **Du, J. (Jinrui)** 35:17  
Thank you.  
Umm.

 **Marcos Malosetti Zunin** 35:26  
But do you have experience with with what kind of data?  
I was mostly medical type of data or.

 **Du, J. (Jinrui)** 35:36  
Do we have more?  
Any experiences in medic being data we we uh.  
I myself worked.  
Well, that was a PNG.  
So retail data, and then we shall works with financial data.

 **Marcos Malosetti Zunin** 35:49  
Hmm.

 **Du, J. (Jinrui)** 35:54  
Or a mixed up she was a doctor, so I guess he will.

 **Marcos Malosetti Zunin** 35:59  
Thanks.

 **Du, J. (Jinrui)** 35:59  
Medical there Sonja is a mathematician.

 **Marcos Malosetti Zunin** 36:04  
Yeah.

 **Kanbar, M. (Mikdad)** 36:04  
Yeah, but I think the master program is more designed into like the health data overall, it's more like affected by the lumc.

 **Marcos Malosetti Zunin** 36:11  
Health, yeah.

 **Kanbar, M. (Mikdad)** 36:14  
And all examples are coming from there.

 **Marcos Malosetti Zunin** 36:17  
Yeah.  
Yes.  
So that's why I'm asking that because for example, I mean the big difference here is that actually we are working with big families.  
Yeah.  
And not like in in, in medical studies where you have small families and and and and you might have also big studies based on twins for example.  
Well, here we have a lot of those, but not twins of two, but maybe of hundreds.  
So that's a bit of A and that's where the the the in the statistical models, it makes a lot of sense to accommodate their their their relatedness between individuals that are in the population universe study.

 **Du, J. (Jinrui)** 36:54  
That sounds fascinating.

 **Marcos Malosetti Zunin** 36:55  
Umm.

 **Kanbar, M. (Mikdad)** 36:59  
So what would you say?  
Like it's more of a high dimensional data that you are working with.

 **Marcos Malosetti Zunin** 37:04  
Yes.  
Yeah.  
And say in the genetic space is it's huge and that's why I also so the the the, the the number of combinations that you can get it's infinite if you like it because how the the genetics of the traces is is composed and you're able to just sample a very tiny part of it and that's the part that actually you put in the field to to test and in a few rounds of selection then you have to decide which is good which is not.

 **Kanbar, M. (Mikdad)** 37:23  
But.  
Yeah. Understand.  
So I would suggest, let's conclude then.  
Thank you very much for being here.

 **Marcos Malosetti Zunin** 37:49  
No.

 **Kanbar, M. (Mikdad)** 37:49  
It's always great pleasure.

 **Marcos Malosetti Zunin** 37:52  
You're welcome.

 **Li, X. (Xiang)** 37:52  
OK.

 **Kanbar, M. (Mikdad)** 37:53  
Yeah, maybe we can do that in the future when they thank you very much.

 **Du, J. (Jinrui)** 37:53  
Thank you, doctor.

 **Song, J. (Jia)** 37:54  
Thank you.

 **Li, X. (Xiang)** 37:57  
Thank you very much.

 **Marcos Malosetti Zunin** 37:57  
You're welcome.  
Yes, have a good better day.

 **Song, J. (Jia)** 37:59  
Thank you.

 **Kanbar, M. (Mikdad)** 37:59  
Have a nice day.

 **Marcos Malosetti Zunin** 38:02  
Bye bye bye bye.

 **Kanbar, M. (Mikdad)** 38:02  
A nice day. Bye.

 **Du, J. (Jinrui)** 38:03  
Bye bye.

 **Song, J. (Jia)** 38:03  
I have to say bye.  
Thank you.

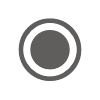
 **Marcos Malosetti Zunin** left the meeting

 **Li, X. (Xiang)** left the meeting

 **Du, J. (Jinrui)** left the meeting

 **Song, J. (Jia)** 38:15  
OK.

 **Kanbar, M. (Mikdad)** left the meeting

 **Du, J. (Jinrui)** stopped transcription