1:42

Alexey

**This week, we'll talk about machine learning in marketing. We have a special guest today, Juan. Juan is a Berlin-based mathematician and data scientist. He is interested in statistical learning, time series analysis, Bayesian and geometric methods in data analysis. Welcome, Juan.**

2:01

Juan

Yeah, thank you very much. I'm very happy to be here. Thanks again for the invitation.

# Juan’s background

2:08

Alexey

**You recently gave a talk at PyData Berlin and I thought that the talk was amazing. I unfortunately wasn't able to attend the talk because they didn't let me in. The talk was already full and I couldn't get in. But I really wanted to talk to you and to invite her here, so thanks a lot for agreeing. This is an amazing topic. I'm very happy to talk about this topic with you today. But before we go into the main topic, which is machine learning in marketing, let's start with your background. Can you tell us about your career journey so far?**

2:47

Juan

Yeah, of course. Before I jump in, I just wanted to say that the video of the PyData talk is live, so you can go check it out. It's online already. So, I'm originally from Colombia. I came to Berlin around 10 years ago to pursue my studies in mathematics. I joined Humboldt University. I did my Master’s and PhD in an area which has actually nothing to do with data science, per se. It was on geometric analysis, which was very interesting. It's something that I wanted to do just for the sake of doing research, especially because I really liked geometry. After spending some time in academia, I decided to do something else.

My first position about data was at TD Reply, which is a marketing consultancy. It was quite nice, because this first experience exposed me to different types of projects and clients in various industries, and it also kind of gave me the business point of view. Because, again, data scientists is not just math and code, it's also about how to make this useful for people to improve their businesses. That was quite fun and took almost three years, but then I decided to move into a product company, because we were essentially doing the more risky projects and proof of concepts for the client, and then we delivered that for in-house development. But I was kind of missing this product development part and that's why I joined HelloFresh.

As probably we'll talk about in a bit, during my experience at TD, I did a lot of time series analysis. I joined HelloFresh to support the forecasting team and that was quite fun, especially because it happened during COVID. Doing forecasts during COVID was definitely challenging, and definitely interesting. We couldn’t, of course, rely on standard methods, so a lot of new techniques and tricks had to be applied. But after that, I really wanted to come back to marketing, again, from a product perspective. For around eight or nine months I've been working at Wolt, where I'm part of the marketing tech team leading the data science projects in the marketing domain.

5:09

Alexey

**Can you tell us a few words about geometric analysis? What is that?**

5:14

Juan

Yeah, sure. Geometric analysis is trying to understand topological invariants of twisted surfaces. I was especially working with surfaces with corners, so to say. But if you see the surface, you can see, for example, if they have corners, if they have holes, etc. But if you have these very high up dimensions, you cannot probably see that and you want to detect these through integrals or through matrices.

One of the ways of understanding it is something like wanting to hear the shape of a drum. What I mean is, if I give you an operator on a manifold and I compute the Eigenvalues – the spectrum – can I detect some geometry? And you can partially detect that. For example, if you take these Eigenvalues, you can see things like what the dimension is, or what the volume is. Yeah, so it’s this game to try to detect global properties through more analytical methods.

6:17

Alexey

**Quite unrelated to marketing, isn't it?**

6:22

Juan

[chuckles] Yeah. But I mean, the core is linear algebra. Linear algebra is definitely a core part of both worlds all the same.

6:29

Alexey

**Yeah, I know this talk is about marketing. But I'm just curious, are there applications of geometric analysis in the day-to-day work of people? As regular people, can we see applications of this somewhere?**

6:48

Juan

As I said, the core components – essentially, linear algebra – are always there. Something where I’m particularly interested in this Bayesian inference approach, which are these kinds of samplers to get a sample from the posterior distribution, actually rely a lot on geometric properties. So actually, people design these samplers based on geometry and it's been quite fun to see how all of these techniques of the remaining geometry and Hamiltonian dynamics can actually be the tool to create these samplers. So it's kind of a far connection, but it exists. It's very interesting.

# Typical problems in marketing that are solved with ML

7:31

Alexey

**Interesting. I did not know about that. Not that I know much about Bayesian inference anyways, but I also didn't know that there is any connection to geometry there. Let's go back to marketing. Hopefully, I do know a few things there. Not a lot, that's why we have you here. So, can you tell us what the typical problems are that we solve with machine learning in marketing?**

7:57

Juan

Yeah, I think this is a very interesting question because there's by no means a complete answer that I can give, just because there are many subfields. On the one hand, the most common one I can think of is about how to optimize media spend – to do better targeting to users. Of course, you want to see which targets to use, like sending personalized messages and so on. You also want to prevent a churn. For that, of course, you have historical data, and you probably have some early regressors that could be a predictor for that, and then you can take action upon these results.

But you can also do things, for example, using NLP and text mining through social listening – that's something that I did in the past – to try to see, for example, how people talk about your brand or about certain campaigns in social media. So you see what the sentiment is, what the subtopics are, and if the campaign’s intention was actually reflected in how people comment on that. So, it's quite huge. Maybe a couple of them that I've been working on at the moment, are on the one hand side, on the user acquisition side, which is how to better use our monitor to efficiently push our marketing activities. And that is somehow related, of course, with the attribution model.

You want to understand the flow – how does the euro you’ve spent work to bring new customers. On the other part, you have retention. Once you have your customers, you want to make sure that they're engaged with the product. For that we can do the churn prevention model, or as I talked about at the PyData talk, uplift modeling to use not just prevent prediction, but actually to prevent churn through tailored incentives.

9:57

Alexey

**So the main ones are, as you said, user acquisition – meaning we want to get new users. Then once we’ve got the users, we want to keep them, or detect if they're about to leave us and somehow prevent this. The talk that you gave was about detecting this, right?**

10:16

Juan

Yeah, exactly.

# Attribution model

10:18

Alexey

**You also mentioned the attribution model. As I understand it, when we try to acquire a user, there are multiple ways of doing this – we can show a commercial on TV, we can put a banner on the street, or we can go to Facebook and show an ad there. There are many, many different options, and the goal of attribution models is to understand how effective each channel is. A user came into our platform – where did this person come from? Right?**

10:52

Juan

Yeah. I think there are two parts. On the attribution side, you spend some euro in different parts – it could be a TV ad or a Facebook ad – and then someone reduced it. So the first thing is to connect what was the incentive, or trigger, for this user to come in? Of course, this is not unique and that's the fun part. Because if you see a TV campaign, you probably won't react immediately, but maybe after a while – maybe that same day – you will Google for that and then through the detailed tracking, you could say, in principle, attribute that to Google. But then this will underestimate the effect of the TV ad.

So in the attribution part, which is also connected to measurement, you want to detect this connection. After you have done so, based on certain assumptions, then you want to optimize that because you cannot simply keep putting more money in the same marketing channels because they saturate. We have seen that in practice. Otherwise the strategy would be super simple – we just keep pushing money into the campaign [chuckles] but we know that it doesn't work like this. So yeah, it's a little bit in that direction.

12:17

Alexey

**You’re saying if I start advertising something on Facebook now, then in a month, I will see fewer users coming in from Facebook. Then I would need to go to a different channel?**

12:29

Juan

Yeah, especially because – let's say there's an audience that is available on Facebook. You can try to reach them, but at some point, the efficiency of each euro that you put on this channel is not going to increase. You really just saturate. But, of course, there are many components. For example time – if you do it close to Christmas or close to summer, this also makes a huge impact. So it's not just the euro, per se, but also the introduction of many features.

# Media Mix Model – detecting uplift and channel saturation

13:03

Alexey

**I'm especially curious about TV. You said – you run a commercial on TV, but then users see the brand, so they may recognize the brand and then the next time they Google something, and they see, “food delivery” and then they click on this. So how do you know that this user actually first learned about you through TV and not through Google? Is it even possible?**

13:36

Juan

Yeah, this is a tough problem. There’s a statistical technique to try to infer that. Actually, this is connected to what's called a “media mix model”. Let me take a step back. Before we had all of the cookies and tracking, marketers actually used these statistical models to try to estimate the uplift, at a channel level. Then the cookie industry came in, you had semi-deterministic tracking for users, so that went away a bit. But TV is a channel on which many companies spend a lot of money.

Sometimes they don't know what the efficiency is, they just know that it works. So now that new privacy measures are coming, for example, the iOS changing – in iOS 14.5 – then all of these methods are coming back to life. And they can work pretty well, but of course, they are statistical methods, so it's hard to say if they work perfectly. I guess one of the main tools for that is, of course, A/B testing or geolocation experiments – we can talk about the details later – but it's possible through statistical methods to measure this uplift.

14:53

Alexey

**What is uplifting in this case? What do we actually measure?**

14:58

Juan

Yeah. There are two ways of doing this – let's say level. You can do that holistically, because the marketing funnel is rather complex. In this case, what you do is a regression model. That's kind of the core of the media mix model, on which you put a target variable, such as conversions, and you would try to explain that through all channels as external regressors. The tricky part is that if you put the raw data as it is, it will probably not work for TV, or for other channels as well, because there are two effects that get mixed in this regression-like relation. On the one hand, there's a saturation effect, which means that, as we said, that it’s not linear.

So it's not really a correlation that you're trying to find. In addition, there's a carryover effect. And a carryover effect means that you measure something – you show an ad today and probably not all people that saw the ad will react on the same day, but a week, perhaps. In this media mix model, what you want to do is to couple saturation transformations and its carryover effects – which are sometimes called ad stock transformation – and you put that into the mix in this kind of regressional setting. Of course, you can control for seasonality, you can go fancy and go with time varying coefficients, because the efficiency of marketing can change over time. But that's kind of more on the holistic level.

Coming back to your question about uplift, it's really the coefficient of this regression. Whereas if you want to do this at the campaign level, where maybe you don't have all of this data and all of this mix, where you want to detect an uplift, then you can go through a slightly different approach, which is the so-called cultural impact. What you do here is train a time series model for your KPI of interest (let's say conversion) by controlling by seasonality and maybe other regressors (other types of media spend). Then you generate a prediction, assuming that there was no campaign, and then you have the true values of your KPI, and then you will attribute this uplift – again, this is a big key – in that period of the TV campaign.

17:30

Alexey

**Maybe I didn't understand correctly, so just to make sure I did – uplift in the first case, when we just look at the contribution of each of the channels. We have a regression model, so the target variable here is conversion – let's say somebody registered in an app or somewhere, or downloaded an app (some action, let’s say registration) and then there are multiple channels that lead to this registration. First, the user could have seen an ad on TV, then maybe they could have heard this on the radio, or in Google, or in Facebook, or… there could be many channels.**

**So you have all these channels, which are the features – the regressors – in this model. Then the target is one of the users. Right? Then you train this model, you look at the coefficients, and this gives you each channel’s contribution to the conversion. You see, “Okay. For TV, the coefficient efficiency is two times more than for radio. Okay, TV must be two times more effective.”**

18:45

Juan

Yeah, in a sense. I think that's the core. There are two things I would like to add. On the one hand side is that the raw impression state or cost data that you put into the model would not be enough. So you need to put this saturation and ad stock effect, which actually have some hyperparameters that you will learn from the data. So you would actually like to learn from the data when these channels saturate. The other is that, in these media mix models, like you said, there are direct and indirect effects. So what you usually do is not have just one regression model, but have a couple of them to model different touch points.

If your target variable is conversions, then you have a model where TV is included. But then you have yet another regression model on which your target variable is, let's say, Google search. And then you have TV as a regressor for that. Then you kind of do an average to see the combined effect. So it's actually a sequence of regression models.

19:48

Alexey

**There’s another thing that I'm really curious about. When it comes to Google or Facebook – you have tracking – you know that this user came from this channel – but when it comes to TV, you don't really know about that. So do you have another model that predicts if this user was exposed to a TV commercial? Or how does it work?**

20:09

Juan

Yeah. You don't do this at the user level, but you usually aggregate daily or weekly granularity. You kind of have a pool of where all of the users are aggregated and the agency that manages the TV data can give you cost, which is how much they spend. And they also have a way of estimating the audience, which would be the impression. So this is what you actually use. It's a time series model, so to say – you have the time component that is important – and you have weekly or daily granularity.

20:44

Alexey

**This is called the “media mix model,” right?**

20:48

Juan

Yeah.

# Changes to privacy regulations and its effect on user tracking

20:49

Alexey

**You also mentioned that we have all these things that track us every time we click on an ad, a cookie – or some identifier of each of us – is somehow saved in the system and we have access to this. But you mentioned that there are changes in some privacy regulations that are coming soon and my understanding is that this kind of tracking will not be possible in the future. Is that right?**

21:19

Juan

Yeah, it already happened actually. I think it was last year on iOS 14.5, for example. In your iPhone, you can actually refuse to share that data with Apple. So what Apple actually reports is not at the user level, but an aggregate report. In that regard, these types of statistical models are not truly affected by this and I believe that this is going to continue to happen. Privacy is going to make it so that the business statistical models which work on aggregate data will be the tool the marketers will need to use because the deterministic way is probably not going to work anymore.

22:04

Alexey

**But you will still know if somebody came from Facebook or not, right? You just don't know if this user maybe visited some other website.**

22:16

Juan

I think you don't even know the aggregate number. You probably cannot identify the user. The report will say “10 users came from iOS.” But you don't know which one.

22:31

Alexey

**Okay. That makes the modeling more complex, right?**

22:38

Juan

Yeah. But again, if you think about TV, you don't have that either. So that's the key component – again, I think the media mix model is not really about the model. It is, but it's really about which data you can find. Because you’re already finding good TV spend data – you need to have a common granularity and that's a big part of the project, the data collection.

# User retention and churn prevention

23:04

Alexey

**Okay. So we do this, we understand how effective each marketing channel is, and then we can decide whether to spend some money in this channel or not. We also should keep in mind the saturation as you mentioned, and then another area was optimizing how much money was spent – what are the budgets are, right? So then, we acquired the user and now our goal is to try to keep this user as long as possible in our application or on our website. So what are the things, what are the models, or what are the problems that we're solving there when it comes to retention?**

23:43

Juan

Retention actually depends on the type of business. If you have a contract-based business then you have a well-defined notion of churn. Typically, this is a classification problem on which you have certain features that will probably explain or indicate why this user churned. And if you think about a classification model that outputs, let's say, a number between zero and one, then you can run them on your customer base and set a threshold, “If it's more than 0.7, I will send an email,” or something like that. This is kind of a very high level view.

In other businesses, like for example Wolt, there's no definition of churn because you could order today and order tomorrow, but then go on a vacation for a month. That doesn't mean you churned, that means that you're probably just not active. So, in this case, there are other types of techniques – there are many of them. Some techniques that I have really been looking into are more the probabilistic type of model, where you try to simulate this non-contractual behavior and try to estimate the probability of being active as a function of time. So it’s not a typical classification problem, but again, that depends on the business model.

25:17

Alexey

**I guess here there is no definition of churn because it's an app or it's a registration, right? So unless a user deletes their account, you don't know whether they deleted an app or just stopped using it, right?**

25:32

Juan

Exactly. As you [cross-talk]

25:34

Alexey

**You cannot track deletion, right?**

25:37

Juan

Yeah, but also by that time, it's already too late. Just to give you an example, the whole idea is to model the purchase frequency. There are customers, for example, that order every Sunday, and there are customers that order every day. So the customer that orders every day stops for four days, there's probably something wrong and then you probably want to react. That's an example. But if a customer that just orders every Sunday and stops for four days, it's normal – it's expected. It's going to be more of a concern if this user doesn't order for a month, so to say. So it still tries to find this kind of sweet point on which you detect that there's something weird from the user panel. And of course, you want to learn this from historical data.

26:24

Alexey

**Well, what if I am from a different category of users? I order when I just feel like it – when I don't feel like cooking. So there is really no pattern for that – one day I might order and maybe I will for the next day, but then maybe for a month or two, I will not order anything. But then because, let's say, I go to work and want to eat out, but then I'm lazy or I have a meeting so I just order out again. You described users who order every Sunday, you described users who order every day – and then there are people like me, who only order sometimes. Do you have a different approach to each of these segments of users?**

27:05

Juan

Yeah. Of course, you will have to have external awareness. So, for example, I would imagine that you order more in winter than summer. [cross-talk]

27:16

Alexey

**[chuckles] I don’t know, I don’t collect the data. [laughs]**

27:19

Juan

Let's say that maybe it’s just because you're less incentivized to go outside and maybe just call and order. It depends, yeah. I'm just trying to tell you the fact that you can, of course, add seasonality features. It depends on the customer as well. So again, as you do in a churn-like problem, where you have a binary variable, you will try to see from a look-alike approach, if you can detect some signal. Of course, it's never going to be perfect, but it's at least something to make sure that we target the right users at the right time. Of course you don't want to get emails every day. It’s super annoying.

27:58

Alexey

**And practically, how is it implemented? Do you have a different model for each segment? Or do you have one model for all the users? Or is it so business-specific that every business needs to do it differently?**

28:14

Juan

I think it depends on the business. But you can think about this SLR regression-like problem, where you can just add external regressors into that or where the output probability will be a function of this, for example, segment or external regressor.

28:31

Alexey

**Usually, in this case, let's say you detect that this user is about to churn. I used a different app (I will not say which app it is) but I used to use it, but then I stopped because there is a competitor that I like more. They started sending me pushes, so they detected that I'm not active. And these pushes are so annoying that they just want to delete the app. I guess my question is, do you also need to take the cost of push into account? Because maybe I didn't delete the app yet. But with these push notifications, you kind of annoy me, so I go and delete it.**

29:13

Juan

Yeah. [chuckles] I'm also very annoyed by these emails. I think these are two different problems in the sense that on the one hand, you want to have a model that predicts the probability that you're active. But then you need to do something else to efficiently target the users that you can actually recover. Because if you are gone, you're gone – why do I need to waste or save money by sending you emails? That costs, right?

In that regard, this is where uplift modeling comes in, where you really want to learn which users are the ones that are useful to target based on historical data. Again, that's what I presented at PyData. And this should be built on top of the churn prediction because we don't want just to predict – we want to prevent. Of course, the output rates are something that companies have to take into consideration. The strategy of just sending emails and hoping it works is a little bit too naive.

30:26

Alexey

**So you also need to be selective. If the model says, “This user is not active anymore,” then you need to see, “Okay, how hopeless is this user?” If the user is hopeless, you don't bother, because the user is long gone.**

30:43

Juan

Yeah, exactly.

# A/B testing to detect uplift

30:45

Alexey

**I guess the factors you use here are like how often the user uses the app and what kind of patterns there are, right?**

30:54

Juan

Yeah. But these uplift models actually need an A/B test. So what you actually need – the training data on this optimal length actually are coming from a trace control split. You do the trace control split, you measure the uplift and then you try to detect signals. Because the problem is that you cannot send and not send an email to a user – that’s what you would ideally like to measure, but you cannot. So what you try to do is find similar users, such as that in the control group you don't send anything and in the treatment you send. And then by comparing the uplift of these two, you can estimate. So if one of them did convert again with the treatment, then you know that this type of users are the ones that you probably want to target. But if they didn't, for example, that's a hint of the model we have to say, “Okay, maybe this type of user, based on certain features, is not the one that you want to target.”

31:53

Alexey

**Practically, I guess, you take all your user base, you somehow cluster them, segment them. Then in each segment, you split them into two groups, A and B. And then you think, “Okay, let's take this segment and then we will send a push or an email. To this group, we will not send anything and we will see how many of them will actually return.” This is how you measure the effectiveness of the feature. Got it.**

32:23

Juan

Yeah, but I just want to say that in real life data collection, I think, it's the most challenging part, to be completely honest with you. Because the models are kind of classical machine learning models, but of course the marketing department would like to just push a lot and to do these control experiments in a way that you know that the only thing that is different is the treatment. It's hard. You don't want to have compounding effects, like different regions, for example – or something happens in a city and another thing didn't happen in that city. So it's tricky.

32:57

Alexey

**You said marketing is pushing – marketing wants to send emails to everyone? So why even bother splitting with A and B? Just send it to everyone and see. [chuckles]**

33:07

Juan

I'm just saying that doing experiments is a commitment that everyone should have in the company. It's not like the crazy data scientists are just trying to do it, but sync across everyone. For example, we also need to make sure that the treatments that we send in the training phase are actually going to be consistent in the feature. Because if I send an incentive without a voucher in my training period, and in my test, or my real-life experiment where I'm going to apply that voucher, then it's not really that consistent.

33:45

Alexey

**Then I guess you can also take a segment and your A/B test would be – to one group you send an email with a voucher, to another you send without a voucher. Because sending the voucher also has some cost, right?**

34:00

Juan

Yeah, yeah, exactly.

34:01

Alexey

**Then you see how much revenue it actually generates in each segment at the end? Right? So does it make sense to send vouchers? Or maybe how large the voucher should be.**

34:12

Juan

Yeah. And I think just to make it even harder – let's say you want to optimize for long term retention. If you just offer a voucher and this person uses the voucher and then doesn't come again, it's a big question whether this was useful or not, right? You really want to make sure that there's long-term engagement and not a short-term effect just driven by incentives.

34:39

Alexey

**But also [chuckles] there could be long-term engagement driven by incentives. There is an app that I use for fast grocery delivery and the only reason they use it is because there is free delivery and they give a 10 euro discount when it's over a certain threshold. The moment they stop doing this, I'll just go to a different app.**

35:02

Juan

Yeah. I think every problem is really trying to understand the customer that we have. Yeah, this diversity makes everything trickier but fun.

# Statistical approach vs machine learning (setting a benchmark)

35:15

Alexey

**We have an interesting question, “Which approach is more efficient – statistical approach or machine learning?”**

35:24

Juan

I mean, I don't have a clear difference between these two. I would say that you should always go with a baseline that maybe neither of those and have that as a benchmark. So I wouldn't jump into these techniques unless it's necessary, because these problems are surprisingly hard. If you have the right data, you might actually get away with a simple rule. Things become a little bit more complex if you don't have the data available. So keep it simple.[chuckles]

36:01

Alexey

**So what could be a good baseline – a good benchmark – for churn prediction?**

36:07

Juan

For example, if they're active last week.

36:10

Alexey

**Okay. That's pretty simple.**

36:14

Juan

Yeah. Whatever you do, this is just the first example that came to my mind. It has to be, because if not, then just use that.

36:24

Alexey

**What do you think the differences are? Because I'm not completely sure. What are the differences here between statistical and machine learning approaches? To me, they look kind of the same. I guess maybe machine learning is like when you train XGBoost and statistics is when you train linear regression? Or is it about tests?**

36:45

Juan

I don't know. I don't have strong opinions about this. For me, it's just methods to solve a specific problem. I do believe that, for example in the media mix model, it’s really about doing a very good linear regression. And in practice, that's actually hard.

# Does retraining MMM models often improve efficiency?

37:05

Alexey

**Yeah, we have a question about this MMM model. I think they mean the media mix model. How often do you train these MMM models, and are there any significant gains in performance, if you retrain them weekly, for example?**

37:22

Juan

Usually, if you think about measuring offline campaigns, you don't have these every week or every day. Probably, what you typically do is to have a good baseline, and maybe do it maybe every month or every two months, that depends on the direct granularity. Of course, the digital channels will keep going, but the strategy is often to just go on and off because it is quite expensive. Of course, we really try to automate as much as possible, like data transformation, data collection, things like this. But retraining on a daily level is not going to bring any value.

# Attribution model baselines

38:11

Alexey

**We talked about a good baseline for churn prediction. What are good baselines here for attribution models?**

38:22

Juan

Again, this is really about the data that we have, because in an ideal situation, the attribution model would be deterministic and then you shouldn't have to model anything. But for example, in the iOS case, you really want to attribute that at the user level – there should be a way of splitting this report into individual users. So I guess the easiest one is just to distribute that uniformly, but the other better methods may be to use look-alike approaches to do so.

38:58

Alexey

**By uniformly you mean, you just assume that every channel is…**

39:03

Juan

That's a little bit tricky, actually. It's not that simple. In an ideal case, for example, the report is “Okay, there are 100 users” and then you have a way of detecting – you don't have per se which ones they are, but you have a subset of “100 are coming from this channel and 100 from another.” You don't know which ones, so then you kind of randomly assign, just so that the report makes sense. But of course, you cannot trace that back. It's tricky.

# Choosing a decay rate for channels (Bayesian linear regression)

39:41

Alexey

**There is another question from Sebasis, which is probably also about this MMM model. Or it's related to saturation, I think. The question is, “How do you choose the decay rate for each channel? And what's the approach that you follow?”**

40:00

Juan

Yeah. Actually, you don't need to choose that. You will actually learn it from the data. The technique that you're using, to be more concrete, is Bayesian linear regression. And again, this Bayesian approach allows you to plug these types of transformations in a nice way and you can actually learn that from the data. The challenge, of course, is that you might not have enough data or you could over-parameterize your model just because you don't have enough data points. And this is where these Bayesian techniques on which you use the priors to shrink the coefficients based on the domain knowledge, for example, or certain heuristics which can help you a lot. So ideally, you could learn this from the data.

# Learning resource suggestions

40:46

Alexey

**Is there any good resource on learning all these things? We talked about the media mix model, we learned about this technique that you just mentioned, Bayesian linear regressions, uplift modeling, churn prediction – is there a good book or course or something that talks in detail about all these machine learning methods, or general data methods in marketing?**

41:14

Juan

I guess there are many resources online. I should say that they're kind of all over the place. So just a little disclaimer, I have a little blog where I try to run some simulations, so that could be maybe a nice place to start. But there are many blogs online about the subject. I have found a conceptually interesting book that is called Introduction to Algorithmic Marketing, which is available online. It gives a very nice overview of the marketing domain – it talks about customer lifetime value, efficiency measurement through MMM, and they go beyond. They also have a nice GitHub repo on which they also have some experiments. I found that reference quite interesting.

# Bayesian approach vs Frequentist approach

42:06

Alexey

**Yeah, thank you. I see a question from Amin. We talked about Bayesian linear regression, and the question from Amin is, “Do you use the Bayesian approach for building your statistical models, or are you more into the frequentist approach?”**

42:22

Juan

I really like the Bayesian approach, because on the one hand, at least for me, it's easier to understand – it’s a little bit more transparent. I know p values can be understood, but I just find it slightly a bit more transparent. And actually, it gives a lot of flexibility. So as I mentioned before, you can also, of course, try to do this with maximum likelihood estimation. But the fact that you can encode business knowledge in your priors is something that comes very handy if you have small data. So it's just a very convenient approach. I use it, but not just because it's fun or cool, but just because in some situations, it does show a big advantage in this specific statistical method.

43:16

Alexey

**You probably talk often (or sometimes) to your colleagues from other companies who also work on marketing. Do you see any preference from your colleagues towards Bayesian methods in general, or is it 50/50, or maybe frequentist is more popular?**

43:36

Juan

I think people working on MMM (marketing mix model), most of them work with Bayesian methods, just because of the flexibility it provides. So in that regard, I think it's very popular. But for other applications, for example, for churn prevention, you'll probably try to use a maximum likelihood estimation or tactical machine learning model, just because you really want to aim for accuracy, and also the scale and the data set are typically much bigger.

44:17

Alexey

**I guess, when explainability and ability to use the business knowledge, the business knowledge is more important, so you go with Bayesian. But when you care more about accuracy, then you go with XGBoost or something.**

44:36

Juan

Yeah, I still believe – again, this is some early experiments that I have been doing – but one of the benefits of Bayesian modeling as well is that you can have this hierarchical structure, which in some sense allows you to solve the cohort problem. So what happens if you have a new cohort – you're doing that at cohort level – and you can pull information from categories.

In this case, I think even if you're just interested in prediction, or accuracy, it could actually be very useful. Again, the problem is about speed and performance, but I think the people working in probabilistic programming are really working hard and making a lot of great progress on scaling these methods so that they run more efficiently.

45:29

Alexey

**Funny, you said that the Bayesian approach is easier to understand, but you probably mean that it's easier to understand the output and then explain it, right? Because to me, every time I try to understand how Bayesian methods work, I see integrals all over the place, and I have some mental block in my head, perhaps because I didn't study geometric analysis. [chuckles] Or maybe it was for some other reason. But to me, Bayesian methods are more complex. If I really want to understand how they work, then I need to go through all these mathematics. That's why maybe I'm not into Bayesian methods much, simply because I don't understand how they work. [cross-talk]**

46:12

Juan

I totally get it. But I don't think it's the fact that you didn’t study geometric analysis. [chuckles] You definitely don't need a PhD in geometric analysis to understand this approach. I need to be very honest, there's a great reference, which is a book called Statistical Rethinking. The author of this book provides online lectures. It's a beautiful book. It’s a very complete book on Bayesian analysis, when you don't see integrals. It’s just intuition and simulations. It's really beautiful and I strongly recommend it. Honestly, I read through the math, I read through the integrals, but it was just by going through this specific book and its use of lectures that helped me grasp it. Then everything became quite transparent. So it's quite popular among Bayesian practitioners. It's a book that I strongly recommend.

47:26

Alexey

**I also heard about another book, I think it's called Think Bias. I think I attempted to read it. I don't remember a lot of formulas there. Have you heard about this book?**

47:42

Juan

Yeah, I heard. But Statistical Rethinking kept me busy for a year. [laughs] So I did everything on it. But of course, there are many other references. But yeah, try it out and let me know because it's really pleasant to read. Maybe you don't have the time – because I didn't have the time at the moment – also go through the lectures. It's also quite insightful.

# Suggestions for creating a marketing department

48:06

Alexey

**Oh, I see we don't have a lot of time left. But there is a question I really wanted to ask you. Let's say – I work at a startup and we just started building a team. There is some product, we have a brand, but we don't have a marketing department yet. We want to start doing this. Maybe there is a person who runs some campaigns on Facebook, but we're mostly in the dark. Now we heard maybe from this talk or some other talk, that data science is helpful – machine learning is helpful for marketing and we want to start doing this. What would you suggest – how should we approach doing this?**

48:47

Juan

Yeah. Of course, there should be a business problem. The problems should be clear. You want to be more efficient with respect to the marketing spend. Everything I talked about today relies on a good data foundation. You have different channels, you have Facebook API, you have Google, and they have different format, different generalities. I would strongly recommend to devote some time to structure the data, before jumping into any marketing – just doing data integrations of the API, for example, designing a data model in the sense of data warehousing, and making sure that the data quality is “good enough” because, of course, it's not never going to be perfect. And just by doing this and looking to the data, the data should guide the models and the techniques to be used. Again, without data, it's really, really hard. So spend some time building up the marketing tech infrastructure to have reliable data.

49:58

Alexey

**So from what I understood – we shouldn't first think about, “Oh, let's hire data scientists and let them figure out how to best spend our marketing money.” First, we should invest in infrastructure, which means hiring a data engineer, I guess, and a data analyst who would work together. The data engineer would build the foundation and then the analyst would actually look at the data and try to make sense of this data. Maybe it could be even marketing analysts, right? So an analyst who specializes in marketing. Then, together, they will build the foundation – they will understand how things work. So let's say we have that. What would be the next steps? Are we ready to bring in data scientists or not yet?**

50:50

Juan

I mean, this definition is a little bit ambiguous. Because if there was an analyst working in this type of data integration and KPI modeling, I'm pretty sure that that person can definitely do some of the fundamentals of the problems that I described. Because, again, there is so of course – but if that person already exists in the company, they would probably do a much better job starting with the baseline model than some external data scientist that’s trying to get cool models into new data. So there's really a lot of domain knowledge here.

Just to give you a concrete example, I work in a truly cross-functional team and I need to work closely with the engineers and data analysts. Because the media mix model connects with the attribution, and then we need to refine our attribution model, and really find the KPI they want to model. So, in a nutshell, it’s not that there's a specific point in time where you need to bring a PhD with geometric analysis by no means necessary. I think just by having a good data foundation, domain knowledge, and a little bit of statistics and linear algebra, you can actually do a lot of interesting things.

52:13

Alexey

**So I guess the most important thing here is domain knowledge, which trumps everything else. A good analyst who knows data well can probably pull together a Python script for doing simple modeling. Right?**

52:29

Juan

Yeah. Because, of course, if you want to do product analysis, let's say you want to apply a churn model, you probably need a little bit more help. It's better to go with “Okay, who is going to set up your Airflow server? Who is going to maintain that?” and so on. That becomes a little bit more tricky. But at least in a very, very early stage, you really need to stop going blind in your marketing spend, but maybe try to start off with some reporting, and some common sense all the same.

52:57

Alexey

**I guess, if we want to have – not data science, but in general – if we want to start this marketing function in our data organization, the first good use case would probably be to spend our money on marketing better, more effectively. That could be a good use case. And the methods that we talked about, like attribution, would be what applies to that. So let's say we want to acquire as many users as possible, we should think about “What is the most effective channel where we should put more money?” And then we also should keep in mind all the things you mentioned, like about channel saturation.**

53:37

Juan

Yeah, maybe to add something to add on top of that – it's also key to define which KPIs you care about, so you can optimize in respect to that. Is it conversions? Which type of conversion? Because you can register today and use the app today, or in seven days? Or do you care more than short term? So defining what you want to optimize for, by looking for the data that you have in place, it's also an important step. You need to see what you want to improve.

54:10

Alexey

**Yeah, interesting. And how do we decide if retention is more important than user acquisition? Who makes these decisions?**

54:22

Juan

Yeah, I think this is really strategic. There should be a vision, right? Of course, there’s value in acquiring customers, but something that I truly believe is that – no matter what you do marketing-wise, if your product is not solving the users’ problem or helping users, then you're just burning money. It's important to focus on marketing, but I think it's also key to make sure that the product is actually going to be the best tool because it's really about the product that drives who is going to join and if they are going to be engaged. Because no matter how many emails or vouchers I send you, if the product is bad, and it's buggy, you're probably not going to use it. So I would also say to focus on product development.

# Most challenging problems in marketing

55:12

Alexey

**So retention in this case is not only about having a good churn model, but it’s also having a good product. A product people want to use. If it’s buggy, if it crashes, then why do I need it? In your opinion, what are the most challenging problems in marketing?**

55:34

Juan

I think I told you – there are many that I keep thinking about and reading a lot about, which is about offline channels and media efficiency. I think these MMM models are quite good on paper and they work beautifully in simulations. But when you need to put this into practice, it’s quite challenging. It's quite fun, because it's hard to find a template where you can “just run it” because it really depends on the data. Often you don't have the data available, so you need to find proxies, or maybe try to do an experiment, or maybe use previous experiments to adjust your priors.

So these are definitely a few problems that I believe require a lot of creativity. Because to be completely honest, I'm not really driven by training fancy models, or super big models that require a lot of computational power. For me, it's about solving problems that require new ideas. So if it's not going to be a Bayesian linear regression, okay, what do I need? I think there's still a lot of room for new techniques to optimize media spend.

# The importance of knowing marketing domain knowledge for data scientists

57:02

Alexey

**How important do you think it is to know marketing for data scientists – if somebody wants to work in marketing? We talked about different terms like funnels, conversions, CTR – all these things. I guess for somebody who wants to go into marketing them, here's all these abbreviations, all these words – that can be quite challenging, right? So in your opinion, how important is having the general knowledge of marketing?**

57:39

Juan

Yeah. So this is something that I had to learn by heart, but the marketing managers are your best allies. I assure you that, even if you have good data, and if you have good knowledge of machine learning, if you're working by yourself, without talking with the marketing department or the marketing manager, the project is going to fail. Because needs change, requirements change, and it's really about the strategy and the plan. So what if you optimize for a channel that they are going to stop using? It doesn't make any sense.

So it can be a little bit tricky, but the marketing team is your stakeholders, and you need to have very transparent and continuous communication with them. It can be quite fun. They have a lot of knowledge that your model actually wants. So it's super, super important. Actually, for me, coming from academia, I was a little bit bored about just talking with mathematicians. Talking with people from different backgrounds can make things a little bit more fun.

# Juan’s blog and other learning resources

58:48

Alexey

**Yeah, thank you. There is a question about your blog, and I did a quick Google search. The blog is Juanito Orduz? Sorry, I cannot pronounce that. Can you please say it?**

59:06

Juan

Yeah, it’s juanitoorduz.github.io. I can share it on Twitter as well.

59:12

Alexey

**I just shared the link. We will also include the link in the description. I think you also mentioned some resources, like some books. The first one was Introduction to Algorithmic Marketing. The other one was statistic… something related to rethinking? Statistical Rethinking. Did you recommend anything else?**

59:46

Juan

There are many resources online. There's a very nice talk. Let me try to look for it. It was a PyData talk that was about churn prevention. I can share it.

60:00

Alexey

**Is it your talk? [chuckles]**

60:01

Juan

No, no, no, no. It's more like a holistic point of view that goes from churn, uplift, and optimization. I think it was PyData 19. If you put churn prevention PyData 2019, you'll probably find it.

# Finding Juan online

60:22

Alexey

**Maybe if you find it later, send us a link and we will put it up. So what's the best way to find you on the internet?**

60:32

Juan

You can find me on the blog. You can also find me on GitHub. You can find my email quite easily. So if you ever have a question, or anything, whether it’s on this or other topics or on data science, just drop me a line.

60:48

Alexey

**And Twitter, I guess, is also another way to find you. Okay, thanks a lot. Thanks for joining us today. Thanks for sharing your experience with us, for telling us about marketing. And thanks, everyone, as well for joining us today, for asking questions, for watching us. I guess that's all from us today.**

61:11

Juan

Alright. Thank you. Thank you very much for the invitation. It was a pleasure.

61:13

Alexey

**Have a great weekend.**