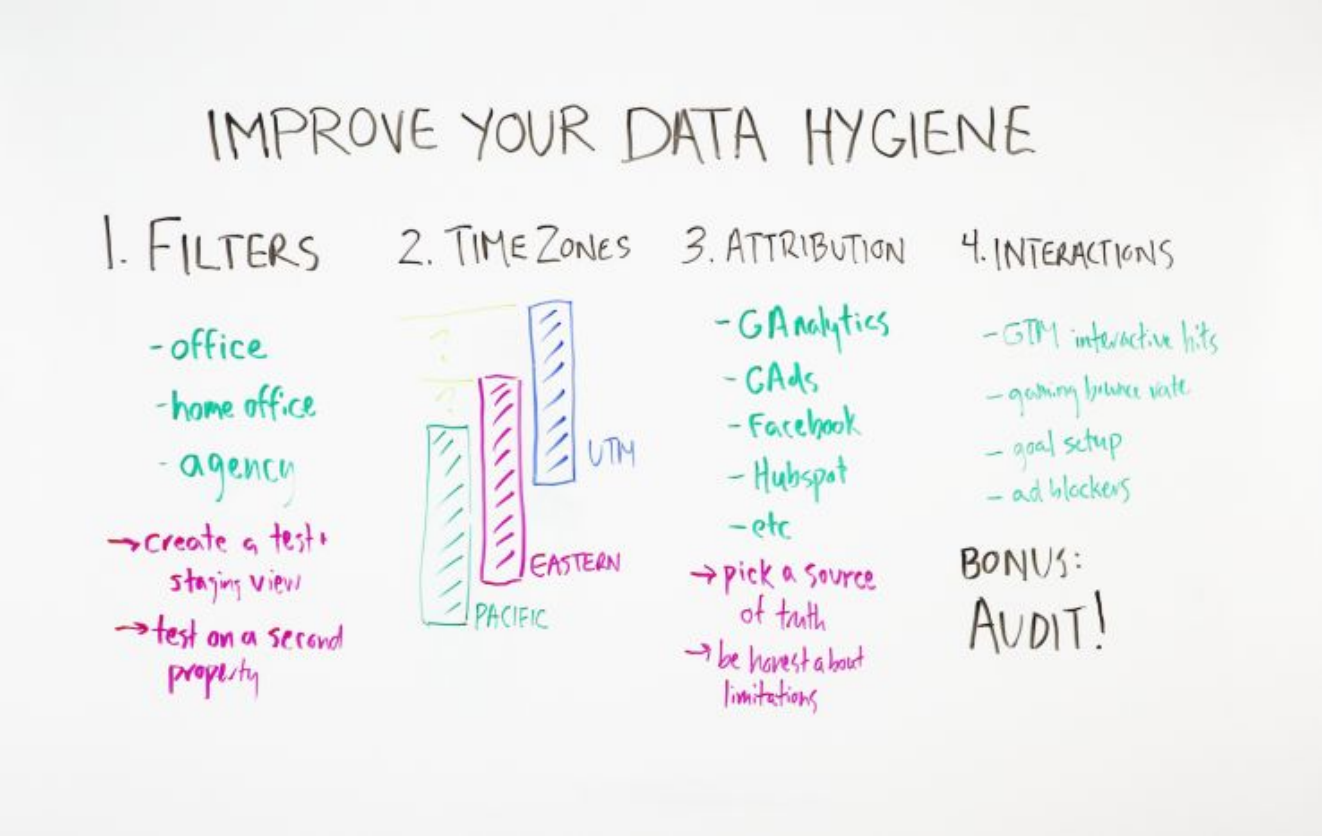
4 Ways to Improve Your Data Hygiene – Whiteboard Friday

We base so much of our livelihood on good data, but managing that data properly is a task in and of itself. In this week's Whiteboard Friday, Dana DiTomaso shares why you need to keep your data clean and some of the top things to watch out for.



Click on the whiteboard image above to open a high resolution version in a new tab!

Video Transcription

Hi. My name is Dana DiTomaso. I am President and partner at Kick Point. We're a digital marketing agency, based in the frozen north of Edmonton, Alberta. So today I'm going to be talking to you about data hygiene.

What I mean by that is the stuff that we see every single time

we start working with a new client this stuff is always messed up. Sometimes it’s one of these four things. Sometimes it’s all four, or sometimes there are extra things. So I’m going to cover this stuff today in the hopes that perhaps the next time we get a profile from someone it is not quite as bad, or if you look at these things and see how bad it is, definitely start sitting down and cleaning this stuff up.

1. Filters

So what we’re going to start with first are filters. By filters, I’m talking about analytics here, specifically Google Analytics. When go you into the admin of Google Analytics, there’s a section called Filters. There’s a section on the left, which is all the filters for everything in that account, and then there’s a section for each view for filters. Filters help you exclude or include specific traffic based on a set of parameters.

Filter out office, home office, and agency traffic

So usually what we’ll find is one Analytics property for your website, and it has one view, which is all website data which is the default that Analytics gives you, but then there are no filters, which means that you’re not excluding things like office traffic, your internal people visiting the website, or home office. If you have a bunch of people who work from home, get their IP addresses, exclude them from this because you don’t necessarily want your internal traffic mucking up things like conversions, especially if you’re doing stuff like checking your own forms.

You haven’t had a lead in a while and maybe you fill out the form to make sure it’s working. You don’t want that coming in as a conversion and then screwing up your data, especially if you’re a low-volume website. If you have a million hits a day,

then maybe this isn't a problem for you. But if you're like the rest of us and don't necessarily have that much traffic, something like this can be a big problem in terms of the volume of traffic you see. Then agency traffic as well.

So agencies, please make sure that you're filtering out your own traffic. Again things like your web developer, some contractor you worked with briefly, really make sure you're filtering out all that stuff because you don't want that polluting your main profile.

Create a test and staging view

The other thing that I recommend is creating what we call a test and staging view. Usually in our Analytics profiles, we'll have three different views. One we call master, and that's the view that has all these filters applied to it.

So you're only seeing the traffic that isn't you. It's the customers, people visiting your website, the real people, not your office people. Then the second view we call test and staging. So this is just your staging server, which is really nice. For example, if you have a different URL for your staging server, which you should, then you can just include that traffic. Then if you're making enhancements to the site or you upgraded your WordPress instance and you want to make sure that your goals are still firing correctly, you can do all that and see that it's working in the test and staging view without polluting your main view.

Test on a second property

That's really helpful. Then the third thing is make sure to test on a second property. This is easy to do with Google Tag Manager. What we'll have set up in most of our Google Tag Manager accounts is we'll have our usual analytics and most of the stuff goes to there. But then if we're testing something new, like say the content consumption metric we started

putting out this summer, then we want to make sure we set up a second Analytics view and we put the test, the new stuff that we're trying over to the second Analytics property, not view.

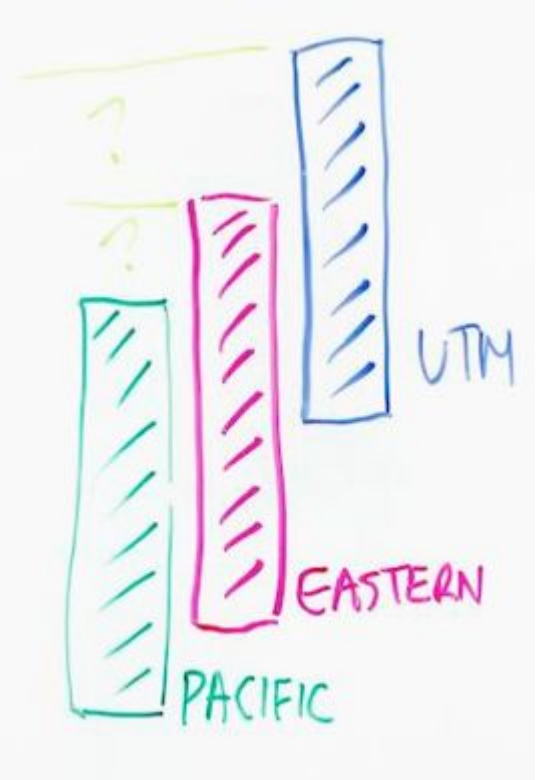
So you have two different Analytics properties. One is your main property. This is where all the regular stuff goes. Then you have a second property, which is where you test things out, and this is really helpful to make sure that you're not going to screw something up accidentally when you're trying out some crazy new thing like content consumption, which can totally happen and has definitely happened as we were testing the product. You don't want to pollute your main data with something different that you're trying out.

So send something to a second property. You do this for websites. You always have a staging and a live. So why wouldn't you do this for your analytics, where you have a staging and a live? So definitely consider setting up a second property.

2. Time zones

The next thing that we have a lot of problems with are time zones. Here's what happens.

Let's say your website, basic install of WordPress and you didn't change the time zone in WordPress, so it's set to UTM. That's the default in WordPress unless you change it. So now you've got your data for your website saying it's UTM. Then let's say your marketing team is on the East Coast, so they've got all of their tools set to Eastern time. Then your sales team is on the West Coast, so all of their tools are set to Pacific time.



So you can end up with a situation where let's say, for example, you've got a website where you're using a form plugin for WordPress. Then when someone submits a form, it's recorded on your website, but then that data also gets pushed over to your sales CRM. So now your website is saying that this number of leads came in on this day, because it's in UTM mode. Well, the day ended, or it hasn't started yet, and now you've got Eastern, which is when your analytics tools are recording the number of leads.

But then the third wrinkle is then you have Salesforce or HubSpot or whatever your CRM is now recording Pacific time. So that means that you've got this huge gap of who knows when this stuff happened, and your data will never line up. This is incredibly frustrating, especially if you're trying to diagnose why, for example, I'm submitting a form, but I'm not seeing the lead, or if you've got other data hygiene issues, you can't match up the data and that's because you have different time zones.

So definitely check the time zones of every product you use —website, CRM, analytics, ads, all of it. If it has a time

zone, pick one, stick with it. That's your canonical time zone. It will save you so many headaches down the road, trust me.

3. Attribution

The next thing is attribution. Attribution is a whole other lecture in and of itself, beyond what I'm talking about here today.

Different tools have different ways of showing attribution

But what I find frustrating about attribution is that every tool has its own little special way of doing it. Analytics is like the last non-direct click. That's great. Ads says, well, maybe we'll attribute it, maybe we won't. If you went to the site a week ago, maybe we'll call it a view-through conversion. Who knows what they're going to call it? Then Facebook has a completely different attribution window.

You can use a tool, such as Supermetrics, to change the attribution window. But if you don't understand what the default attribution window is in the first place, you're just going to make things harder for yourself. Then there's HubSpot, which says the very first touch is what matters, and so, of course, HubSpot will never agree with Analytics and so on. Every tool has its own little special sauce and how they do attribution. So pick a source of truth.

Pick your source of truth

This is the best thing to do is just say, "You know what? I trust this tool the most." Then that is your source of truth. Do not try to get this source of truth to match up with that source of truth. You will go insane. You do have to make sure that you are at least knowing that things like your time zones

are clear so that's all set.

Be honest about limitations

But then after that, really it's just making sure that you're being honest about your limitations.

Know where things are necessarily going to fall down, and that's okay, but at least you've got this source of truth that you at least can trust. That's the most important thing with attribution. Make sure to spend the time and read how each tool handles attribution so when someone comes to you and says, "Well, I see that we got 300 visits from this ad campaign, but in Facebook it says we got 6,000.

Why is that? You have an answer. That might be a little bit of an extreme example, but I mean I've seen weirder things with Facebook attribution versus Analytics attribution. I've even talked about stuff like Mixpanel and Kissmetrics. Every tool has its own little special way of recording attributions. It's never the same as anyone else's. We don't have a standard in the industry of how this stuff works, so make sure you understand these pieces.

4. Interactions

Then the last thing are what I call interactions. The biggest thing that I find that people do wrong here is in Google Tag Manager it gives you a lot of rope, which you can hang yourself with if you're not careful.

GTM interactive hits

One of the biggest things is what we call an interactive hit versus a non-interactive hit. So let's say in Google Tag Manager you have a scroll depth.

You want to see how far down the page people scroll. At 25%,

50%, 75%, and 100%, it will send off an alert and say this is how far down they scrolled on the page. Well, the thing is that you can also make that interactive. So if somebody scrolls down the page 25%, you can say, well, that's an interactive hit, which means that person is no longer bounced, because it's counting an interaction, which for your setup might be great.

Gaming bounce rate

But what I've seen are unscrupulous agencies who come in and say if the person scrolls 2% of the way down the page, now that's an interactive hit. Suddenly the client's bounce rate goes down from say 80% to 3%, and they think, "Wow, this agency is amazing." They're not amazing. They're lying. This is where Google Tag Manager can really manipulate your bounce rate. So be careful when you're using interactive hits.

Absolutely, maybe it's totally fair that if someone is reading your content, they might just read that one page and then hit the back button and go back out. It's totally fair to use something like scroll depth or a certain piece of the content entering the user's view port, that that would be interactive. But that doesn't mean that everything should be interactive. So just dial it back on the interactions that you're using, or at least make smart decisions about the interactions that you choose to use. So you can game your bounce rate for that.

Goal setup

Then goal setup as well, that's a big problem. A lot of people by default maybe they have destination goals set up in Analytics because they don't know how to set up event-based goals. But what we find happens is by destination goal, I mean you filled out the form, you got to a thank you page, and you're recording views of that thank you page as goals, which yes, that's one way to do it.

But the problem is that a lot of people, who aren't super great at interneting, will bookmark that page or they'll keep coming back to it again and again because maybe you put some really useful information on your thank you page, which is what you should do, except that means that people keep visiting it again and again without actually filling out the form. So now your conversion rate is all messed up because you're basing it on destination, not on the actual action of the form being submitted.

So be careful on how you set up goals, because that can also really game the way you're looking at your data.

Ad blockers

Ad blockers could be anywhere from 2% to 10% of your audience depending upon how technically sophisticated your visitors are. So you'll end up in situations where you have a form fill, you have no corresponding visit to match with that form fill.

It just goes into an attribution black hole. But they did fill out the form, so at least you got their data, but you have no idea where they came from. Again, that's going to be okay. So definitely think about the percentage of your visitors, based on you and your audience, who probably have an ad blocker installed and make sure you're comfortable with that level of error in your data. That's just the internet, and ad blockers are getting more and more popular.

Stuff like Apple is changing the way that they do tracking. So definitely make sure that you understand these pieces and you're really thinking about that when you're looking at your data. Again, these numbers may never 100% match up. That's okay. You can't measure everything. Sorry.

Bonus: Audit!

Then the last thing I really want you to think about – this is the bonus tip – audit regularly.

So at least once a year, go through all the different stuff that I've covered in this video and make sure that nothing has changed or updated, you don't have some secret, exciting new tracking code that somebody added in and then forgot because you were trying out a trial of this product and you tossed it on, and it's been running for a year even though the trial expired nine months ago. So definitely make sure that you're running the stuff that you should be running and doing an audit at least on an yearly basis.

If you're busy and you have a lot of different visitors to your website, it's a pretty high-volume property, maybe monthly or quarterly would be a better interval, but at least once a year go through and make sure that everything that's there is supposed to be there, because that will save you headaches when you look at trying to compare year-over-year and realize that something horrible has been going on for the last nine months and all of your data is trash. We really don't want to have that happen.

So I hope these tips are helpful. Get to know your data a little bit better. It will like you for it. Thanks.

Video transcription by Speechpad.com

http://tracking.feedpress.it/link/9375/11073214?utm_medium=feed&utm_source=feedpress.me&utm_campaign=Feed%3A+mozblog