**Diagnostic analytics — how to conduct a root-cause analysis**

**Proactively delivering real insights into metrics changes**

*Note: The term root-cause analysis is commonly used in IT and data engineering as the process to identify root causes of faults or problems. This article focuses on diagnostic analytics to understand the drivers of business metrics changes.*

Why is revenue down? Why did the conversion rate spike? Why is the average order value flat?

Depending on your industry and business goals, your key metrics might be different. But if you care about answering “How do I improve my key metrics?”, you need to first understand why they changed first.

I see many teams struggling to answer to identify the root causes behind business metric changes. Depending on their diagnostic analytics maturity level, the reasons vary (see [diagnostic analytics gap](https://towardsdatascience.com/the-diagnostic-analytics-gap-3f9d0a44e8f8)).

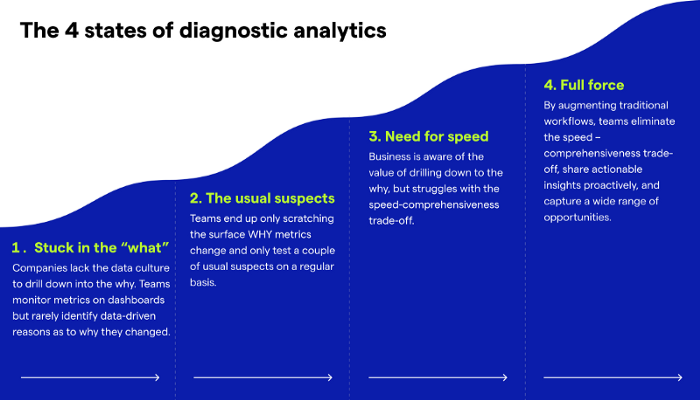
**Why is this relevant?**

To be able to come up with recommendations to improve performance, teams first need to understand the changes happening in their metrics. Without looking under the hood it is not possible to find true insights -not just observations- for teams to act on (see the article “[How to deliver true data-driven insights](https://towardsdatascience.com/how-to-deliver-real-data-driven-insights-e94f2b386c18)”).

Ideally, they should do this regularly as opposed to investigating when a sharp decline or spike is observed in metrics. Otherwise, teams will get stuck in reactive analytics loops which could be quite costly due to delayed insights and actions.

**Status quo**

Based on conversations and collaborations with dozens of data & business teams, one can group diagnostic analytics maturity under 4 states, depending on the approach to diagnostic analytics, analysis thoroughness, and time to insight.



(States of Diagnostic Analytics — Image by Kausa)

**Stuck in the what**

Teams are mainly describing what’s happening (e.g., the metric is going down) and connecting the dots with high-level qualitative facts (i.e., events that happened in the business like website updates). The process is unstructured and not data-driven. They are mostly in firefighting mode, looking into the ad hoc requests coming from the business teams.

*Outcome*: Untapped potential that teams are not even aware of, weakened data culture

**The usual suspects (i.e., “business-led” approach)**

Teams have a very high bias towards only testing the usual suspects or the hypotheses proposed by the business teams. While it is useful to receive context and direction from the business units, teams should go beyond just testing these hypotheses. This can introduce significant bias and make teams overlook valuable opportunities.

*Outcome*: Significant bias in the decision-making process, leading to missed opportunities. Lack of real insights.

**Need for speed**

Teams are in a more mature state of diagnostic analytics and they recognize the value of drilling down to the why. While they have a more structured approach, performing comprehensive root-cause analysis with existing workflows is too complex and time-consuming. Often they do not get to the insights as fast as the business demands.

*Outcome*: Insights are uncovered too slowly to make a true business impact and improve decision-making.

**Ideal state — the full force**

The full-force diagnostic analytics requires a proactive approach based on an impact-oriented comprehensive root-cause analysis.



(Slide by the author — Reaching full force in diagnostic analytics)

**Fast AND Comprehensive**

Teams use machine learning to augment diagnostic analytics. This way they can test all the possible drivers behind the changes within minutes, without sacrificing speed or comprehensiveness and removing human bias.

**Proactive**

Teams not only look into drastic/surprising changes but keep a pulse of their metrics on a daily/weekly basis depending on the speed of the business. They go into the performance meetings already equipped with the potential drivers for the changes to drive the conversation. Proactively share these insights daily/weekly instead of reacting to questions from the business.

**Impact-oriented**

A small change in one of your largest markets could be still significant but a major change in a tiny subgroup might appear way more prominent unless you are taking true impact/actual contribution into the account. ([see more here](https://towardsdatascience.com/why-analytics-needs-an-impact-driven-approach-43179f3bec4e)).

When looking at a specific subgroup, its contribution to the global metric change has to be derived from 2 effects:

* Change in the given metric for this subgroup
* Change in the subgroup volume (share of this subgroup over the global population)

These two effects can be calculated individually, assuming *ceteris paribus* (all other things being equal). By adding up both effects you get the contribution of this subgroup to the global metric change.

Let’s assume you are looking into Average Order Value (AOV) changes week over week.

The Global Average Order Value has increased from 50€ to 52€.

You want to understand the contribution of each country to global change. Let’s take France as an example:

* ***Change in metric:*** AOV in France increased from 60€ to 70€
* ***Quantity/volume*:** the share of French orders stayed constant at 10%
* ***Actual impact***: derived by the sum of both effects
* **Change in metric:** this increase in the AOV of France orders contributes positively to the global AOV. Assuming *ceteris paribus* (i.e., the size of France would have remained 10%), the effect is (70€-60€)\*10% = 1€
* **Change in volume:** 0 (given that the size remained constant)
* Given both effects, if everything else were to stay the same, the global AOV have moved up 1€ (i.e., this is France's contribution to the global AOV change)

**The best approach to conducting root-cause analysis**

The best practice methodology consists of 3 key steps: narrow down the change to the most impactful subpopulation(s), analyze related/dependent metrics, and connect the dots with actions taken, interactions, and external events. Then it’s critical to ensure those insights are properly communicated and presented to drive actions.

**1) Narrow down the change to the most impactful subpopulations**

Narrow down the change to subpopulations driving most of the change after testing all potentially relevant factors to avoid human bias. This is the part that is humanly impossible to do with manual resources. Even if you are spending days/weeks on a single change, it is highly unlikely that you would look into all the factors and score them according to impact.

Using machine learning / augmented analytics, thousands, even millions of combinations can be checked within minutes and only the significant ones can be scored according to the actual impact explained earlier.

**2) Analyze related/dependent metrics**

After narrowing down the change to the most impactful subpopulations, look at how related/dependent metrics evolved for these subgroups. Oftentimes a metric is a function of other metrics (e.g., ROAS in gaming is a function of expected revenue, cost, and installs). In these cases, having this metrics tree in mind can help further understand why things are changing (e.g., ROAS decrease due to a spike in cost)

**3) Connect the dots**

Collaborate with the business teams to connect the dots and drive actionable insights. Brainstorm on which actions were taken, as well as interactions, events, and external factors that can influence the business metric you’re looking into. Relevant facts can be segmented into 3 groups:

* **Internal** (i.e., actions taken by the company)
* **External and industry-specific** — actions that were taken by the competition (e.g., new marketing campaign) or general trends in the industry (e.g., summer seasonality) that affect the business metric you’re looking
* **External and geographical-specific** — broader actions that affect a specific region (e.g., elections in Brazil)

**4) Present and communicate the insights effectively**

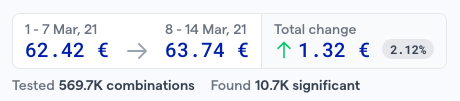
Explain key facts and drivers upfront, stating your hypotheses and quantifying the impact. Develop consistent ways to present results (e.g., waterfall charts, memos like the one below)

**Example**

Let’s go back to the Average Order Value example — AOV

AOV tracks the average dollar amount spent each time a customer places an order on a website or mobile app. Product/Marketplace teams aim to maximize AOV by testing different campaigns and making changes on the website/apps over time.

In this case, AOV increased from 62.4€ to 63.7€ WoW. This is a 2.1% increase.

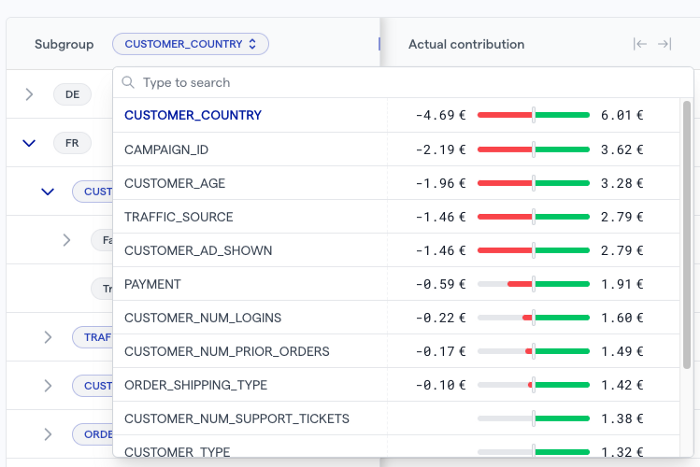


(image by Kausa — metric change snapshot)

As it isn’t a big spike or a drop, it would be pretty likely for many teams to ignore this. At the full-force level, every week key business metrics like AOV are under examination to uncover every opportunity and maximize business impact.

So, where is this 2% coming from?

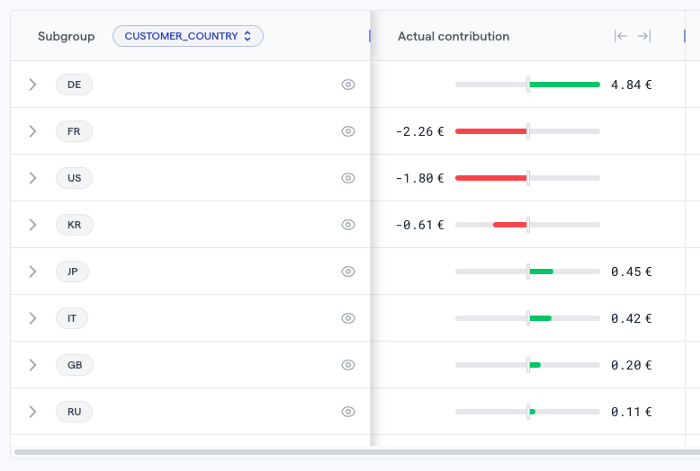
By augmenting the workflow, you could test all the factors affecting AOV — more than 500.000 combinations for this case — and see which drivers have the biggest impact.



(image by Kausa — prioritization of main drivers)

Within seconds, I can tell that country, campaign, and customer age are the biggest impact drivers. And interestingly, there are quite a bit of factors counter-affecting one another and canceling each other out in a high-level view.

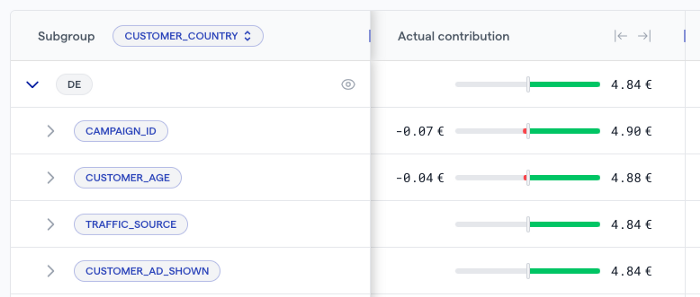
Let’s see which countries are contributing negatively and positively.



(Image by Kausa — actual contribution by country)

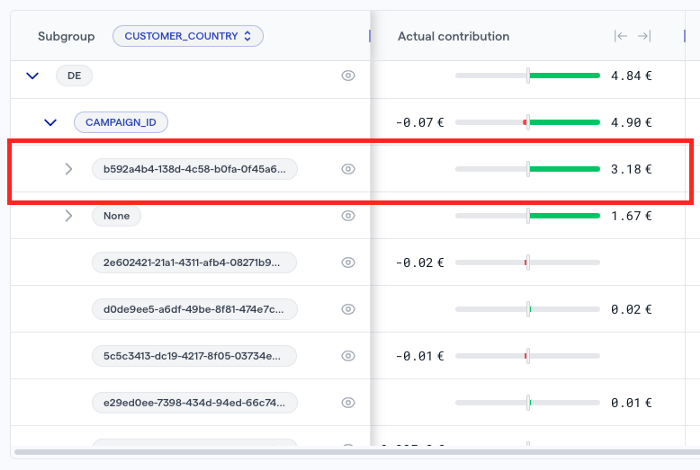
Interestingly, Germany is performing really well while France, the US, and Korea are not so much.

You’re curious why Germany is performing particularly better compared to other countries. So you drill down a bit more to identify that Campaign ID is the subfactor that contributes to the AOV increase in Germany. But which campaign/s?



(Image by Kausa — main drivers prioritized for Germany based on actual contribution)

There is one campaign in Germany that contributes 3.18€ to the overall AOV. Interesting…



(Image by Kausa — Campaigns in Germany prioritized based on actual contribution)

Now you look into related metrics to gain more context. It looks like the marketing team increased the ad spend for this campaign, creating a boost in both order value and order volume. The volume of the subgroup is rather small. You might want to inform the teams we can test in other segments too.



(Image by Kausa — Viewing relevant metrics)

Before you connect with marketing teams, you want to see if you can gather any other insights. Age seems to be an important subfactor. The campaign is performing especially well for the 18–20 age group.



(Image by Kausa — Looking into three dimensions together)

Great! Now you are ready to communicate these findings with the marketing team to gain more business context.

After a quick check-in, you find out that they are testing new creative visuals for this campaign. With this information in hand, marketing teams start testing the campaign in other regions and data teams will be on the lookout to see if the same trend is observed in other countries. And you can allocate time for more interesting data analytics projects instead of spending hours/days on drilling down.

Win-win.

**Bottom line: Impact-driven approach**

Looking at the status quo, there is a big opportunity to improve diagnostic analytics and maximize business value. Currently, the way that diagnostic analytics is handled can sometimes feel like mayhem (from constant firefighting to missed deadlines, etc…). But it is possible to bring logic and structure into this madness. Where to get started? Depends on your starting point but consider the following:

* Evolve your modern data stack to include a decision intelligence/diagnostic analytics platform to be able to augment the workflows
* Start testing all the hypotheses at hand to surface overlooked opportunities
* Score the drivers affecting your metrics based on their actual impact to be able to focus on what truly matters
* Communicate the findings with relevant stakeholders proactively

—

*Thoughts? Reach out to* [*João Sousa*](https://www.linkedin.com/in/joaoantoniosousa/), *Director of Growth at* [*Kausa*](https://www.linkedin.com/company/kausa-ai/)*. Stay tuned for more posts on how to nail diagnostic analytics and increase the value of data & analytics.*