

Applying Analytics to Reduce Node Split Costs

Mohammad Zubair, Sudheer Dharanikota

Unless you know how to use analytics as a tool to solve a business problem – you will get sucked into wasting your time and creating not so useful results. In this blog we show, with an example (get control over the node split costs), the steps involved in solving a complex problem. Note that analytics is only a step in the multiple stages involving in a business problem. Pay close attention to the methodology.

Problem - Sudheer my node split costs are +200% to -200%!

That's how our day in the life as a consultant begins!

Our client is under pressure to identify what is going wrong with the node splits (a network action to alleviate congestion) that is swinging like a full-blown pendulum. We are talking about millions of dollars of discrepancy here.

Taking it little slow – when you budget for a construction project, granted there are many material and labor cost variables, the variance should be +/- 10%. If the variance is 20 times the norm – you got a problem.

So, what do you do now? You have 3 days to figure out because that is all your client have.

Take the right help early to understand the data

Of course, you were given couple of megs of data. There are these costs with all the weird categorization - some you understand (material, in house labor, contract labor related) and some you have no clue (Vendor X APXY1025 – 1230nm). And there are some project related data – where is this project happening, name of the project, miles constructed etc.

You give a panic call to your analytics lead. He nonchalantly says – send me the data, I will take a look at it. Then starts the barrage of questions about the data. Ok.. Ok.. you are hitting the limit. Time to call the guy who actually knows about the bits and pieces. Get help from the domain expert first.

Here are few things to keep in mind in organizing the data with both the experts on –

- Create a dictionary to classify the data you received into reasonable categories
- Normalize the data to compare the apples to apples costs (such as cost/mile)
- Compare or analyze the normalized costs within the same categories

You may ask – when are you going to talk the math! Not so fast sport.

How to apply some of the fancy analytics to solve a real-world cable operator problem? Here we show some of the steps we took to reduce the node split cost by 50%.

What is the problem?

We were asked by a dear friend and an operator – our node split costs are swinging extremes. He gave me the whole weekend (well – 3 days actually, including Friday) to figure out what is going on. He said, “I have a Monday morning meeting, you guys pretend to be smart – solve it”. Yup – we took the challenge and solved the puzzle.

Key Takeaways

- Partner the domain expert and the data scientist together
- Hypothesize before delving into the details
- Apply right analytical tools that are applicable for your problem

Have a hypothesis - You need to narrow down your assessment though intuition



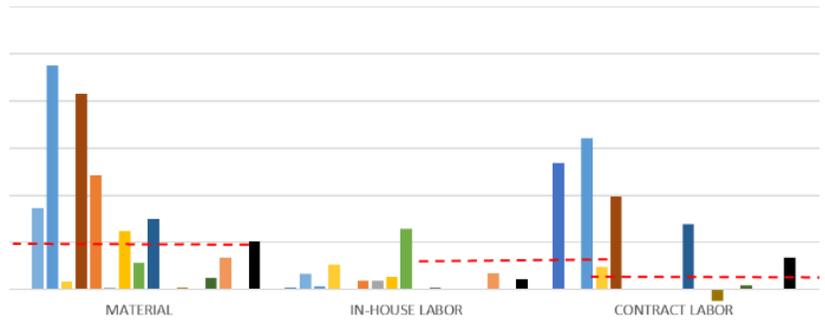
The problems come from different angles – process related, finance related, operations related, architecture related, pure strategy related etc. And you have gazillion ways to skin the cat. But the problem is you have only three days to solve the problem!

You got to start with some hypothesis – are these process related problems, material problems (excess spares, wrong parts, architectural issues etc.) or labor problems (in efficient in-house labor use, excessive errors, improper coding of the work). If you do not narrow them – there goes your time to complete the task.

Now start working in the opposite direction. How can you prove if it is a process problem or a material sparing problem or a labor related problem? How can you get the intuition out of the big (or not so big) data that you are given? Here are some hints –

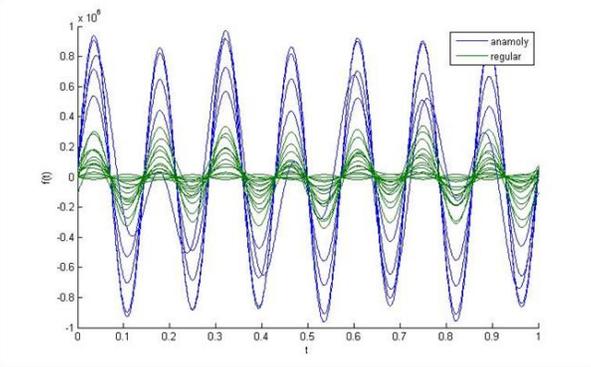


- Start with visualizing the data – plot multiple dimensions against each other
- Take some basic statistical measures against each other for example between different markets
- Understand if there are any outliers in your data under different categories – why is that optics costs high in XYZ market?
- Get the intuition of how can you access different issues – process, material, labor, architecture etc.



Once you have an idea of where you are heading with the data and what are the potential problems - then get into the math.

Look at the major problems - What are your low hanging fruits, major contributors ...

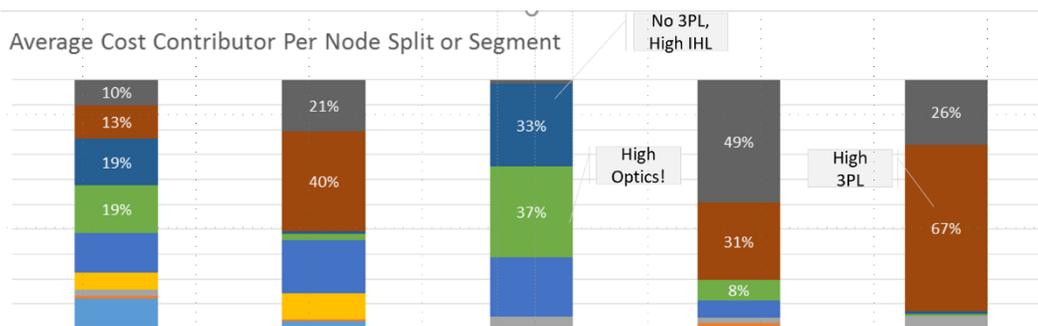


Now you are getting into the geeky stuff.

You have 10+ different variables in the data and you need to figure out what is the relation between them – more importantly which projects are the outliers. If you are talking about two or three variables you can visualize using a nice graph. But the problem is if you have more than three variables how do you show the dependencies. There come multi-variate analytical tools handy. Here are some pointers. You may -

- Reduce your 10 dimensions to two or three for initial visual analysis
- Reduce your 10 to say 6 or 7 principle change contributing dimensions
- Use different visualizations (such as Fourier transform) to show outliers

This is mainly to find the outlying projects (in this case the node splits) – say the top ten offenders. Once the outliers are found – let your hypothesis play to figure out what are the causes for being the offending project. Could be due to exorbitantly higher than normal material costs – because your sparing strategy is not fine-tuned (or not followed). This could be a simple process related correction. Then we get into some serious comparisons such as the costs are outrageous because they are spending too much on the contract labor or excessive expensive in-house labor. You can find some of the expensive trends here.



Then comes the not so low hanging solutions. You may have to go deep into the problems such as architectural and strategic in nature. Our goal here is not to go over them in detail. We will soon in

a detailed white paper. Note that many of our recommendations are straight forward. But the process of getting to this place is where the fun is.

In case you are wondering, yes we did get through the analysis and proposed the areas of correction to our client in three days.



About the Author



Mohammad Zubair is currently a professor of computer science at Old Dominion University. Earlier he worked at IBM T.J. Watson Research Center. His primary research interest is in the application of high-performance computing to large-scale data analytics and scientific computing. In the research area of large-scale data analytics, he has looked at applications of GPU and Hadoop/Apache Spark for: financial risk management, Monte-Carlo based models, and large graph analytics that helps in identifying communities. In the area of data analytics, he teaches two courses: High-Performance Computing and Big Data (Graduate course), and Data Science with Python (Undergraduate/Graduate course). He has received support for his work from NSF, DTIC, DARPA, Jefferson Laboratory, NASA, Los Alamos, AFRL, NRL, JTASC, Sun Microsystems, and IBM Corporation. He has been advising DTS on data analytics for the last four years.

CONTACT INFORMATION

Mohammad Zubair, DTS
Analytics Advisor

Duke Tech Solutions Inc.

+1-919-961-6175

contact@duketechsolutions.com

www.duketechsolutions.com

