

The 10 commandments of Business Driven Data Analytics



If you take the strategic roadmap of any big corporation, I bet you find lines such as *"to exploit our data assets to make better decisions"* or *"to let our data drive are business"* or *"to become a data driven company"*, etc... All good and meaningful... All a must do and a mandatory step to remain competitive and to stay in the market.

Data-driven Business Decisions or **Data-Informed Business Decisions** -to include also the qualitative data- are basically those that are inferred and taken out of the data, backed up by data and verified and justified using data.

It goes hand in hand with **Business-Driven Data Analytics**, which in opposition to mere Data Analytics, is a practice always tied to a business insights need or a business goal. In certain company setups, where the Business Intelligence function remains centralized and the data still stays confined in centralized silos [waiting to be democratized](#), you find a lot of examples where *Data Analytics has become an end in itself*... The results might vary from fancy slides, sexy-looking maps, focus on what looks great and what sells better... but sometimes leaving essential business needs unaddressed and failing at exploiting the real potential in the data.

Data-driven Business Decisions and *Business-Driven Analytics* are *inseparable*, actually the same thing and you cannot have one without the other (like a Yin and Yang).

In this post, we gathered the **top 10 requirements or commandments** you need to observe and "obey" to make sure your Data Analytics is truly "Business Driven":

1. You shall bring your Data where the business knowledge is available

Exactly to where the business decisions are taken, where you have experts dealing with problems on a daily basis... Data can be a life-saver, the vehicle for them to reach their targets! Not doing that is the same as undervaluing your data assets... as *your data is as valuable as the impact of the business decisions you can infer from*.

2. You shall play with data and technologies to generate the rights insights at the core of the business

Nobody gets it right in one run... prototyping, [exploring your data](#), mining relations between data sources, extracting patterns, etc... all that is a *mandatory part* of the data science process and data productization. You need to create a *Lab environment*, where different technologies, algorithms, visualization approaches and all kind of deliverables are tested, prototyped, benchmarked, thrown away in a daily basis (you might get some ideas reading [How to setup a Data Innovation Lab](#))

3. You shall maintain a data catalogue with all your data documented and you shall make your data discoverable

In large corporations, the variety of data available introduces the need for documentation. *Pervasive Business Intelligence* teams might need some data they don't even know about... *Discovery mechanisms* shall be put in place on top of a centralized data catalogue to make the data discoverable and searchable. Standard metadata

shall be made available at data source level (including *time granularity, spatial granularity, labeling and tags, usage statistics, "relevant for", source system or application, how often it gets up-to-dated, etc*). [Linkage](#) -we can see with which other sources a particular source is linked- between data source is ideally also documented, as well as [how to access different data sources](#) with executable example query. Once data has been "discovered", access shall be granted in a similar way an Apple User downloads an app from Apple App Store... a self-service with a sort of contract to be agreed with by the first usage and immediate availability for free ones or after having been granted it (for more restrictive apps).

4. You shall not trust data whose quality has not been checked and documented.

The *quality* and *reliability* of your business decisions is going to be in the best case as good as the quality of the data you used to infer them. That's why you need to be sure you can trust your data, for which the role of a [data quality manager](#) is a big *must*. e.g.: continuously auditing the quality of the information, applying expiring policies, establishing metrics to inform about the recency and trustworthiness of the information, making sure the information remains consistent across reporting systems, that gaps in the data are identified and acted upon, etc... This person is also in charge of maintaining the data catalogue and making it accessible.

5. You shall not use own business taxonomies, you shall keep your standardization efforts centralized.

A [Data Business Taxonomy](#) (broadly speaking, a description of the data at different hierarchical levels) need to be made centrally available (e.g.: with standard product names, standard ways of grouping items, standard ways of describing the campaigns, etc). This taxonomy covers all data standardization requirements company-wide (no department needs to create their own categories). Defining a way of grouping information inside a department might bring you some head-aches in the short term, latest when you share your finding with other departments... and it can go on for years!

6. Your Data shall be provided in a continuous, consistent and (near) real time basis.

One-off data deliveries are better than nothing but there is something called [opportunity window](#) which also applies to the business insights (e.g.: understanding your marketing competitiveness before your more fierce competitor launched a new flagship product line might be only of little help). Hence, the closer in time you get your data, the longer you stay within this opportunity window. [Discontinuous data streams](#) are again better than nothing, but force the business data scientist to apply gaps correction algorithms, interpolation procedures, etc which might extract away important features and require an additional effort that is a non-sense when the real data is available somewhere else in the company. The same apply to externally acquired data sources. [Data consistency](#) should be a given... for example situations where let's say your backend is not responding or off for maintenance and all your orders get queued in some middleware for later processing need to be corrected... otherwise you get a record day but the day before shows an astonishing poor performance... You don't want your business data scientist applying smoothing if you can report consistent data from another source :)

7. You shall test your insight products in an end 2 end environment.

Modern architectures allow for insights to be fed into different systems for automatic [take to action](#). The insights consumer is not always made of flesh and bones, but increasingly often an additional system... On the other hand, your [insights product](#) -or "[intelligence module](#)"- is part of an [ecosystem](#), where it coexists with other components constantly feeding data for [real time decision making](#) and others expecting your intelligence module to provide some outputs. In other words, the result of your business driven data analytics doesn't exist on its own, but as part of an [ecosystem](#) and therefore you need connectivity to all other relevant components (or prototypes of them) from the very beginning (at development stages). It is such an staging environment where your product it is going to prove for business readiness.

8. You shall enforce software development standards to the data application you build

It is not a secret, that the data science discipline has an increasing overlapping with [software engineering](#). No data scientist is complete without [software development skills](#) and best practices. The [first lesson in the Coursera Data Science certification](#) is how to use a version control system (git, GitHub). When it comes down to produce charts, to develop ETL scripts, etc... you better adhere to the software development best practices, such as the already mentioned version control, continuous integration, standard code documentation, project

management tools, [team collaboration tools](#), code generation tools, containerized setup and of course [testing](#) of all kind (unit, component, integration, regression, penetration, performance, etc). The industry is also witnessing that the typical software developer profile is melting with the data scientist profile, as stated by Microsoft Research [in this study](#).

9. You shall be seamlessly engaging your business decision makers

And you do it by showing your progress, discussing your results in regular checkpoints, producing charts, analysis, reports, visuals, click-dummies, etc... Business Driven Data Analytics is it called like that, because *it emerges within a business unit to address a business need*. The process of getting this need addressed is inherently iterative... the early engagement of those whose decisions you are going to support with your insights -business managers- is a key element, working in both directions:

- the business data scientist gets much more business exposure -being pointed out to similar decisions, experts feedback, further relevant data sources, etc- and
- the business manager feels part of the process and is not given a glass ball a few weeks later, and that plays again a essential role in the acceptance.

For that you might require *an additional more stable environment* to let your insights customers give a play... call it pre-prod or call it business playground environment.

10. You shall have a cross-functional combined team ready to scale up also in the business side

Needless to say that Business Driven Data Analytics can only be achieved with mixed cross-functional teams that are capable of scaling up. When you hear about the need for scaling, you immediately think of technical skills, commodity hardware, etc. Business Expertise and business involvement is also "scalable material".

This post didn't pretend to play any evangelizing role... and I'm not a preacher... Actually, if you take each and every aforementioned points, you cannot happen but to acknowledge that they are not right or wrong.. they are just part of the **survival roadmaps** companies need to implement to *maintain* or *increase* their *competitive advantage*.