NEXT LEVEL
DATA ANALYTICS

BREAKTHROUGH SOLUTIONS
TO MEET THE TOUGHEST IT AND
## BUSINESS IMPERATIVES YET

incorta

MOTIVATION
THE NEW DATA IMPERATIVE

The challenges of the past year—economic hardships, government shutdowns, and social distancing — force business teams to think outside the box and embrace new levels of creativity to address the disruption and uncertainty.

Agility and speed of decision-making have never been more important. Business leaders need to move faster, dig deeper, innovate constantly, and be prepared to pivot and transform overnight if they are to survive.

To move forward with confidence, they need the right data–and they need it delivered in away that empowers them to independently discover, explore and act upon the data.
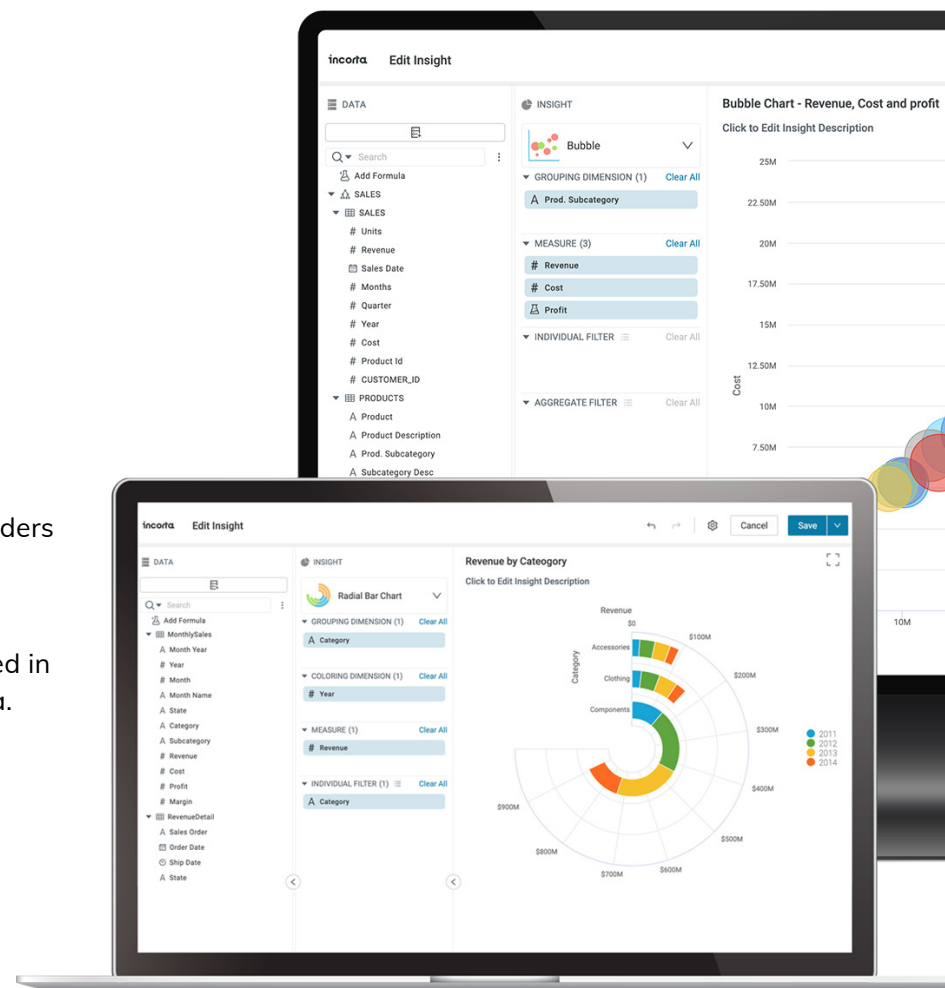
BRIDGING THE DATA GAP

What is holding businesses back from realizing a data-driven agenda?

Despite big investments in data initiatives, the data that business teams need is often not available or delivered in time to make critical business decisions.

There are several fundamental issues:

- **Access to Data.** It takes too long for most organizations to make data available for critical business decisions because traditional ETL and Data Warehousing projects require scarce expert skills like design, engineering, testing, and training.

- **Data Fidelity.** Many analytics systems require data to be reshaped and flattened — changed such that it doesn't resemble the original. Not having access to the original data opens issues of trust and makes it almost impossible to find and correct problems with the data or the business itself.

- **Data Silos.** Gaining access to all of the necessary information sources is so difficult that decision-makers often give up. They rely instead on flawed financial and operational data that is incomplete and often different from what others have gathered.

The good news is that modern, unified data analytics platforms solve most of these problems.

OVERCOMING THE TOOLS GAP

Even with the right data in-hand, business teams struggle because they are saddled with legacy toolsets, complex solutions, or solutions that experts won't touch.

Finding the right balance between ease-of-use, power, and extensibility has been difficult because most buyers of analytics tools were traditionally very technical.

- **Legacy Tools.** Technical staff prefer to use specialized skills to select, configure and stitch solutions together, using existing investments. Unfortunately, this doesn't help business teams become more agile. If anything, it makes them even more dependent on solutions they don't understand.

- **Too Complex.** Data analysis tools are still too complex for many business users without extensive training — if they even attempt to use the tools at all. So they settle for spreadsheets instead. And this means more data copies and not being able to freely navigate all the data or collaborate with their peers. Teams can't be agile if they can't self-service their own data needs.

- **Not Robust.** Some self-service data analysis products are relatively easy to use, but can't grow beyond a departmental solution. They lack the performance and security required by IT departments and aren't attractive to advanced practitioners like data engineers and data scientists.

BUSINESS TAKES THE LEAD

A clear and distinct back-to-basics trend is happening as business leaders are forced to take action.

Teams that were once content to wait for their IT experts to gather, prepare and deliver data no longer have time to wait. They need data now to make the strategic decisions that will ensure business survival.

Non-technical business leaders are now assuming a lead role in new data analysis initiatives. They drive innovation and change in their companies and move faster with more confidence as they become increasingly data-driven.

New, modern data and analytics technologies empower business people to ask questions of their data in real-time. They are gaining the ability to access all data and explore, drill in, and discover insights without having to call up an expert every time they have a new question.

TERMS AND CONCEPTS
SPEAKING ANALYTICS LIKE A PRO

When speaking with vendors and technical people about your data needs, you want to separate modern concepts from legacy techniques that will hold you back.

Extract-Transform-Load (ETL) — This is an old data 'pipeline' concept that says we should (E) gather all the raw data we need[1], then (T) run a process to transform (reshape, combine) the data into a new format that is optimized for the questions we need to answer, then (L) load that data into a data warehouse[2].

Data Warehouse — A type of database that is set up for data analysis (vs. transaction processing) which contains processed and curated data from potentially many different sources. It is meant to be authoritative and may retain large amounts of historical data.

The problem to be aware of with both ETL and Data Warehousing is that you lose all connections to your original data, and worse, you can only find answers to questions that the data model was designed for.

Data Lake — A designated storage area for mostly raw data coming from potentially many different sources. The data is stored as-is, without processing, but in a compressed format that is organized for quick reading. Advances in technology make it possible to run all different kinds of analysis on this data, promoting flexibility and agility.

---

[1] Instead of doing a database query on the source system that might combine, filter and/or aggregate the results before we get them.

[2] A slightly more modern variant of ETL is Extract-Load-Transform (ELT) which defers the transformation step until all the data is in a database or data lake, such that the transformations can be created on an as-needed basis, enabling faster experimentation but requiring a high level of skill and potentially higher computing costs.

BUSINESS INTELLIGENCE VS. ANALYTICS

A couple more terms that are often thrown around interchangeably are Business Intelligence (BI) and Analytics. It is important to understand the differences.

In general, BI is simple reporting that focuses on answering the question, "what happened?" For example, "how much did we sell last month?" "Who are our biggest customers?" BI is often available directly from accounting systems. In large organizations, you may need to pull from many systems for consolidated reporting.

On the other hand, Analytics generally means taking two or more metrics or data streams and combining them together to realize a fresh insight, or show a facet of the business that is not readily visible. For example, "how does geography affect the sales of heavy coats versus socks on a seasonal basis?" In this example, we are combining three separate pieces of data: the geography or region, the sales figures, and the time of year or the season. Analytics focuses on "why" questions and can help spot trends and enable a more forward-looking business approach.

In this eBook, we assume you are looking to make better decisions about the future of your business, so we are talking about Analytics.

DATA BY ANOTHER NAME

As you progress along your analytics journey, you will constantly encounter two broad types of data. An understanding of these terms will help you work with analytics tools and expert practitioners.

In the classic *dimensional model* used by data warehouses, there are two major categories of data that each serve a different analytical purpose:

Dimensions — Sometimes called attributes, dimensions are used to slice the data into meaningful contexts, and represent various facets of the business such as product categories, sales territories, fiscal quarters, organizations, or transaction types. An example would be slicing total revenue by the month and region dimensions.

Measures — Sometimes called facts, measures are numerical values such as currency, quantities, and percentages that are combined with dimensionality to arrive at reportable business metrics. For example, we can report on the percentage of product returns (measure) per month (dimension).

There is one other term that often causes confusion:

Key Performance Indicator — A KPI is just a special case of a metric. Let's say we set a goal to keep monthly product returns under 5%, and we further decide that this is critical for business success. Keeping an eye on KPIs keeps our business healthy. KPIs are often represented as a single number on a dashboard, and this type of formatting is sometimes called a "KPI."

Even if the chosen analytics platform does not utilize a dimensional model, these terms are prevalent in most analytics tools.

ROOT CAUSE ANALYSIS

As you develop your business metrics, numbers that don't look right will sometimes appear. You want the ability to find the source of bad numbers so that problems can be readily identified and solved.

Problems in your metrics will be either due to a data problem, or a business problem.

Data problems can include missing transactions, data that wasn't entered properly, or a problem with data formats. A business problem might be a missing shipment, a systemic shortage of capacity, or a broken business rule that is causing transactions to be coded incorrectly.

Regardless of the type of problem, you want to be able to identify the source of the problem, solve or fix it, and then potentially implement changes to avoid problems in the future.

These are the essentials of Root Cause Analysis. The critical capability that you need is the ability to open up the details behind any metric, "drilling down" all the way to individual detail transactions if necessary. You might need to take action with an individual customer, or there might be a problem with how certain transactions are being transferred to your analytics system.

An analytical system that retains linkages back to all of the underlying detailed data, or even better, a system that performs analysis directly against detailed data, will provide you with the ability to fully explore your data, and find those needles in your data haystack.

HOW TO BEFRIEND IT PEOPLE

At some point in your analytics journey, you will likely need to work with IT people for expert help or to access additional data. Understanding their concerns will help you move your project forward without delays.

IT folks are in a tough bind. On the one hand, they are responsible for how well the organization is using data to make good decisions and drive the business forward. On the other hand, they are held accountable for any security breaches, system downtime, errors and a host of other scary possibilities. It's a thankless job.

Security is important for certain types of data sources. Sometimes systems contain personally identifiable information (PII), and due to privacy laws, IT people are generally not going to entrust this data to you.

There are a couple of ways around this. One way is to ask a data engineer to set up a view3 in the source system that excludes the sensitive data. Another way is to ask IT to set up your analytics system so that it exactly replicates the security present in the source system.

Incorta can do this, and is one of the few analytics platforms that large companies trust with PII data.

Your IT partners will want to help you make your system better, with more data and more accurate calculations, because it will save them time and money over the long run. A system that any business user can operate will save IT from repetitive work. And a system with robust features will be able to grow and evolve with the company, and not become yet another 'legacy' system that they have to clean up or replace later.

---

3   A view is a phantom data table that is derived from the base tables, and may exclude sensitive details. If they create views, make sure the views are at the same level of detail as the base tables, and that they also set up joins.

STEPS TO SUCCESS
ACCESS YOUR RAW DATA

The first step towards agile analytics that will rock your (business) world is to acquire data from all the systems you are using every day. Be ready for some raised eyebrows.

When just starting out, begin with data from systems that you are familiar with. Your department probably 'owns' the data, and so starting within your own domain lets you verify analytical results yourself.

To connect to the data, you'll need to fill in a form with information like a 'system name' or 'connection string', username, password, and other options and settings. The system administrator may just give you this information, or they may want to fill in the information for you.

What is important here is that you gain access to all of the 'raw' or 'unprocessed' data, typically in the form of lots of related or joined data tables[4], and that you can pull updated data any time you want.

Why do you want this? Because business today is crazy, and we are looking for game-changers. We aren't settling for the prepackaged answers that have been curated and blessed by someone else who may not even be in your department.

Don't let anyone talk you into waiting. "Hey, if you can just hang on for a couple of weeks, we can have the data for you all cleaned up." Don't fall for it. Engineers, trying to be helpful, will generally only give you what they think you need. Tell them very politely, but firmly, that you really do want all of the data in its raw form.

[4] Say "We want our data in Third Normal Form, or the closest thing to it" if you want technical people to know you mean business.
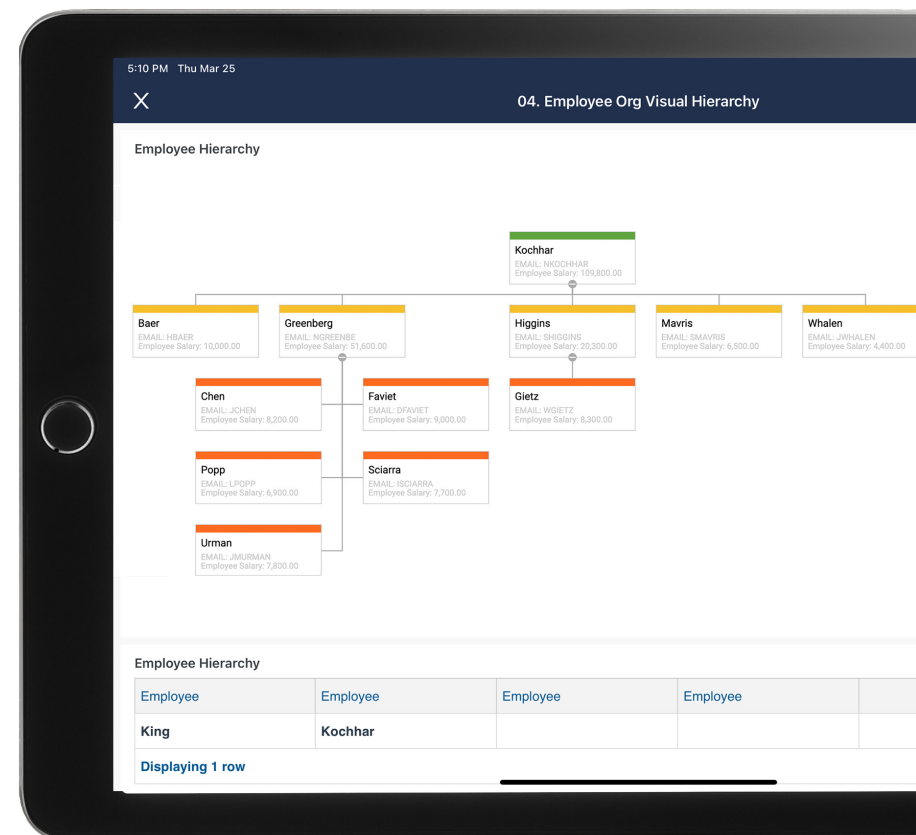
EXPLORING THE DATA

Why analyze data that isn't organized and clean? Because raw, detailed data lends itself to a more free-form iterative approach that prioritizes finding value above careful process.

Traditional Business Intelligence and Analytics projects typically start with a fixed set of questions, and then work to create a perfectly clean and efficient data model to answer those questions. Ah, but what if I come up with a new question? Precisely.

We advocate a more modern, agile approach to analytics, because we can, and because businesses are finding that it delivers results. The approach involves acquiring data as it appears in the source system, and then iteratively pulling out pieces of that data as needed, as opposed to engineering data transformations and an intermediate data model upfront.

"As-is" data will be organized into many tables, usually related (or joined) together. For example, a sales order record is typically joined to a customer record, a product record, a salesperson record and so on to build up a complete picture of the transaction. If the table joins were not automatically read in with the data, you might need to check with your departmental application expert, consult the documentation, or invite your analyst buddy to lunch and nicely ask them to make some sense of it.

These data structures might look intimidating at first, but once you figure it out, you'll have the power and freedom to go anywhere with your data exploration and discovery. Data Scientists understand the power of raw data and as citizen data analysts, we want no less.

Now is a good time to try a few data explorations. As you work, check your results against the source system to be sure you are arriving at the same numbers. Don't wait for perfection. The important thing is to start now, iterate, discover, and grow from there.
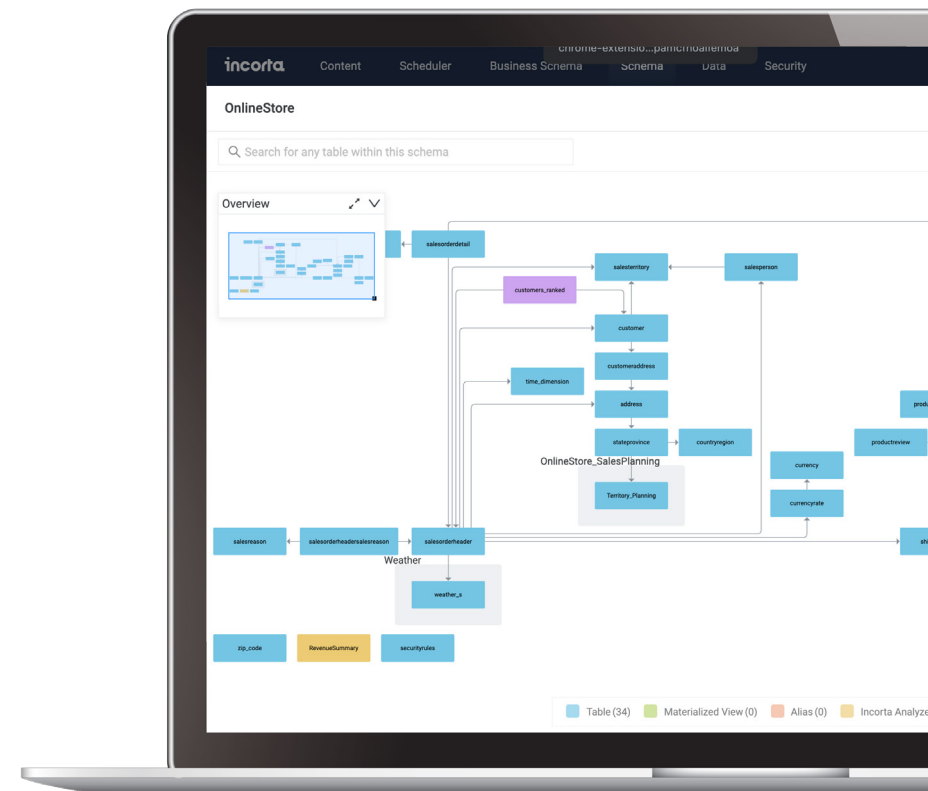
DECIDING WHAT TO MEASURE

How do you decide what to measure? Should you stick to what you are accountable for? Should you use traditional markers of success? Or is there a measurement that everyone else has missed?

Many times you don't have a choice; you are given KPIs, or OKRs, or something else to which you or your department is held accountable. This may even have been why you started looking at the data in the first place.

Domain experts will have ideas based on experience of what to measure to ensure business success and there are sometimes regulations that dictate certain measurements. For example, an insurance company must measure loss reserves, and can also measure various success metrics like the average time to settle a claim.

Also, industry associations are a good source of metrics. For example, if your business involves a supply chain, the Association for Supply Chain Management publishes the SCOR model which includes over 250 standard metrics calculations. Consulting firms such as eCapital Advisors and analyst firms such as Gartner publish industry-specific metrics and benchmarks.

Another interesting twist is in analyzing data from outside your firm. Many companies with interdependent business operations share data. You can evaluate not only your performance, but the performance of your suppliers, vendors, and downstream customers, if appropriate.

However, the best analytical insights are those that shed new light on the business, either to understand new growth opportunities, efficiency patterns or cost savings opportunities. This is where analytics shines, when you combine two, three or even four different factors together to discover something previously unseen.
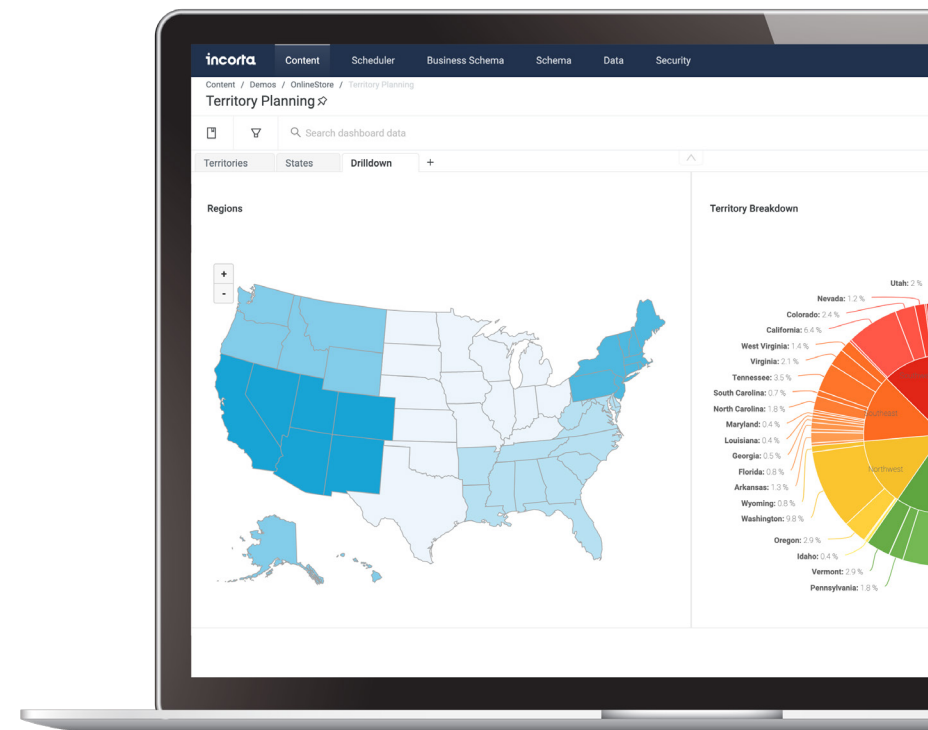
HOW TO ANALYZE DATA

There are many different tools and techniques for analyzing data that go beyond adding up a column of numbers that can lead to new discoveries.

Filtering and Sorting — A basic technique that you will turn to time and time again. As an example, suppose you are analyzing sales data, and wish to focus on only the current year's data, in your sales territory only, and on customers who spend more than $10,000 per month. Then you can sort this data by revenue amount. Could the top customers represent your best upsell opportunities?

Aggregations — This involves computing a sum, count, or average for a particular measure sliced by one or more dimensions that are meaningful to you. For example, it might be useful for us to know which product categories are our best sellers by region. Do men's and women's coats sell equally well in all regions? Are mountain bikes popular in only some states? A Pivot Table, that puts dimensions along the sides and top of a table, and values in the intersecting cells, is very handy for viewing aggregated measures cut by two or more dimensions. A variation of the pivot table is the Heatmap, which shows the hotspots in the table visually.

Trends — Most all transactions have dates, and different parts of the date (the year, month, quarter, day of the week, week of the year, etc.) can be used as a dimension for analysis. For example, what if we want to know the average sales order amount (a measure) by each month (a dimension)? Is it going up or down? Can we detect seasonality using a week-by-week analysis?

Correlations — One of the most powerful tools for analysis is to correlate two or three different values to uncover patterns. Bubble Charts combine three different measures along with one dimension. For example, we can plot product category revenue by margin on the Y-axis, and average discount on the X-axis with the bubble size representing total revenue. This might let us clearly see loss leaders and identify any problems with our pricing strategy.

Above all, don't be afraid to experiment, as trying new combinations of measures and dimensions with different chart types can lead to unexpected insights.
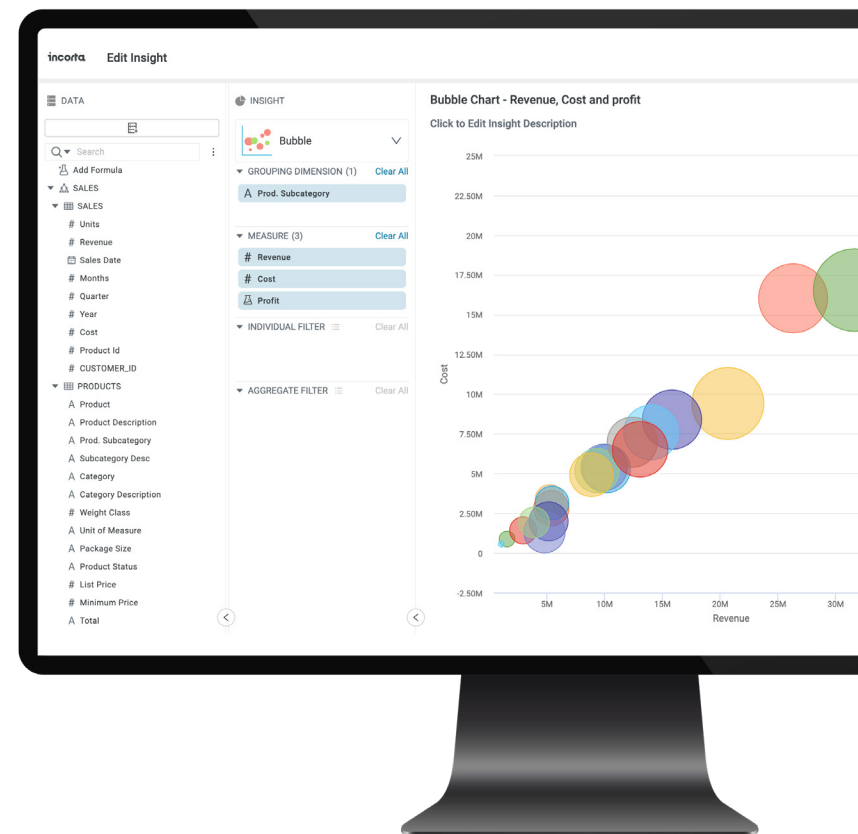
VISUALIZING THE DATA

Carefully choosing your data visualization style can both make it easier to interpret data and cause people to interpret your data in different ways.

You might be familiar with bar charts and pie charts. But there are so many other kinds of visualizations available that can help you tell a story with your data.

We already covered a very powerful kind of visualization, the Bubble Chart, in the last section. The Scatter Chart is similar, but only plots two measures instead of three, and is also great at revealing correlations.

Data that trends over time is often best shown using line charts, while comparisons (for example, sales by category at a fixed point in time) are often best shown using bar or column charts. Note that if your data has 'holes' in it, for example, your monthly data is missing certain months because there were no transactions, then Time Series charts will ensure that the missing values still maintain space in the chart.

Showing the composition of many parts related to a whole is the classic use of a pie chart. But did you know that you can also have multi-level pie charts (sometimes referred to as Pie-Donut or Sunburst charts) that can represent, for example, both categories and sub-category relationships in a single chart?

Other kinds of charts excel at representing geographical data (maps) and chronological data (calendars and Gantt charts) and there are even combination charts that can juxtapose two or more different styles in a single visualization.

The possibilities are almost endless, but the important thing to remember is that different kinds of charts can be used to emphasize exactly the story you want your data to convey.

PUBLISHING YOUR RESULTS

After gathering, analyzing and visualizing your data, the final step is to display your work, either to gain support for your decisions, or to allow others to reach their own conclusions.

Creating an individual chart or data table, especially when you have uncovered an insightful data pattern, can be very gratifying, but it may not be enough to tell the whole story.

A complete presentation of the data may need multiple views or chart types, or the addition of detailed data tables, or even the ability to filter the data to zoom in on particular aspects. Grouping different data visualizations together is the job of a dashboard.

A dashboard typically contains one or more charts, tables, maps, plots, gauges or key metrics. It puts all the data together into a form that people can easily browse. The goal is to arrange all the elements to maximize comprehension.

The ordering and organization of the visuals is important. The layout can either be all on a single page, or across multiple tabbed pages. Tabs can break a larger data analysis apart into smaller pieces that can be sequenced to reveal a data story (sometimes this is called data storytelling.)

The key with dashboards is to keep it simple. There are many advanced products that let you create dazzling infographics, but require talent in the graphic arts to avoid creating a disaster. You want a dashboarding tool that delivers consistently beautiful results without requiring too many choices.

You also want the flexibility to give users the ability to do their own filtering, and drilling down into the data. For example, you might give the dashboard to a group of regional managers — each manager is going to be interested in viewing the metrics for their region, and may want to drill into the details by sales person, by customer, or by product category.
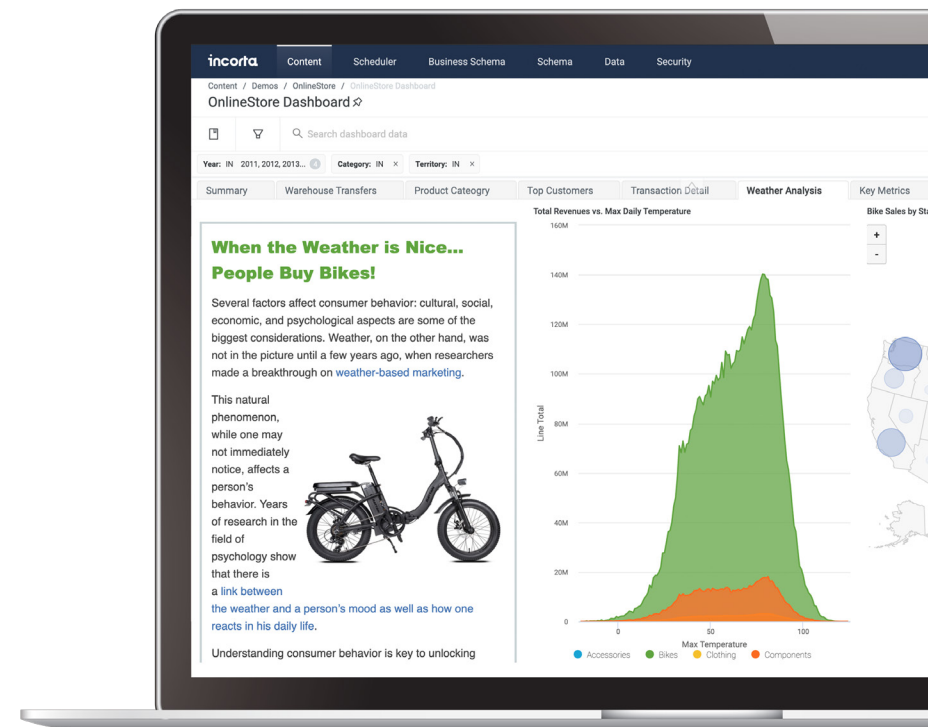
WORKING WITH DATA EXPERTS

At some point, as your skill increases, you may decide that you need to bring advanced practitioners into your project to improve the depth of your analysis.

While our primary message in this eBook is that business people need to take the lead and just get started, there is certainly a time and place for more sophisticated techniques.

Data Analysts are experts at sifting through data to find correlations and identify trends. They typically produce 'Descriptive Analytics' that answer the question: 'What Happened?' They are typically also experts both within their business domain, and with the software platforms used. A data analyst might be able to suggest better ways to capture and utilize the data coming from your systems of record, and might offer valuable ideas for computing metrics and useful KPIs.

Data Scientists are typically experts at creating predictive models and in the correct interpretation of data. They generally produce 'Predictive Analytics' that answer the question: 'What may happen?' They are mathematical experts who are usually skilled at writing advanced data enrichment logic. Data scientists can build experiments to see if it is possible to accurately predict business outcomes or automatically handle routine decisions.

Both data scientists and data analysts prefer to use a platform that isn't under-powered, and will look for certain needed capabilities.

Data Analysts will want the ability to create sophisticated data derivations, statistics, and computations that need to be done in stages and then the results stored. They might use languages like SQL, or Scala to write complex data transformations.

Data Scientists love having access to lots of raw data. They like being able to easily retrieve that data into their tools of choice, or perhaps work inside of an integrated notebook with support for languages like R and Python.

By planning ahead, you can ensure that advanced practitioners will be able to do their work in your chosen platform, and deliver results directly back into the business analytics, saving you on cycle time, aggravation and costs.

CONCLUSION

CONSIDER INCORTA 5 FOR YOUR DATA ANALYTICS

Incorta is a proven solution for IT teams seeking breakthrough performance without the complexity, cost and proprietary lock-in of other solutions. Incorta has evolved into a modern unified data analytics platform with the ease-of-use that business users need to gather, explore, discover and then act on their own initiative, reducing the burden on IT.

The core attributes of Incorta that make it essential for line-of-business teams who want agility without placing limits on growth:

- Access to all corners of the organization with over 200+ data connections easily configurable.

- One platform. The most essential data analytics tooling in a single data experience.

- Easier to use than traditional alternatives; the ideal choice for universal deployment.

- Support agile analytics project methods with access to all original data.

- Can grow into an enterprise-wide solution with security and reliability.

- Attractive to advanced practitioners with integrated notebooks, language support, integration options and support for on-premises, cloud and hybrid deployments.

The approach to business analytics that we have described in this eBook might seem unorthodox, maybe even a little reckless to some, but thousands of people (some with very conservative IT departments) are applying these modern, agile principles to data analytics every day, and achieving incredible business results.

We invite you to prove out the value of Incorta today. Bring us your toughest data challenges and see immediate value.

Get Started.
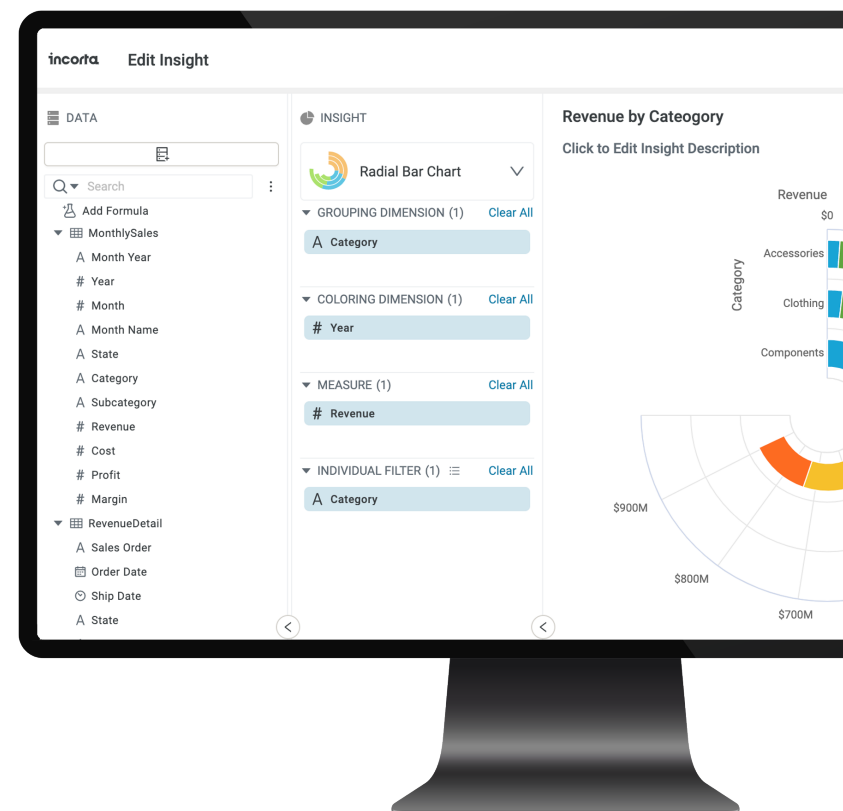
THE DIRECT DATA PLATFORM™

incorta