



# Analytics

- Short Course

Transcript

## ANALYTICS

LESSON 1: UNDERSTANDING DATA.....	3
LESSON 2: ANALYTICS FUNDAMENTALS.....	9
LESSON 3: BIG DATA .....	15
LESSON 4: AI, MACHINE LEARNING, ANALYSIS TYPES, AND FORECASTING.....	20
LESSON 5: REPORTING, VISUALIZATION, AND PRESENTING DATA .....	28

## LESSON 1: UNDERSTANDING DATA

[1.1.1.1] Before we can begin to understand data, we need to define it. Essentially, data is information in raw or unorganized form such as letters, numbers, symbols that refer to, or represent conditions, ideas, or objects.

[1.1.1.2] Data is potentially limitless and it is important to recognize the challenges this poses. Let's think about some of the core challenges related to data. Well, the first challenge is *amount of data* available.

The second challenge is that of *interpretation*. With the abundance of data, we can try and manipulate it the way that we want, and try to get it to tell the story that we want as well. This can lead to falsehoods, which is one of the big issues with analyzing data.

The third challenge is *security*. This is becoming a fundamental issue in terms of privacy. There have been plenty of examples where data has been compromised, and was leaked on to the internet, so it is crucial to think about the way we store and secure data.

The fourth challenge concerns *relevance*. Data, quite frankly, is unequivocally linked to time. With so much data, it can become irrelevant quite quickly. Today, real-time data is actually becoming something that people, particularly consumers, expect, so we need to react very quickly.

The *accuracy* of the data is another key challenge. How accurate do we make the data? And is it capturing all the factors and variables that we want it to?

And then finally, coming back to the point about relevance, data is *time-bound*. It is important to make sure that you use the data quickly, almost in real time, as that is becoming an essential factor in business today. Largely speaking, organization systems often make it difficult to do that.

[1.1.1.3] Now, let's think about some of the more common key misconceptions when thinking about your data. The first misconception is that you need lots and lots of data to do analysis. This comes back to statistics. A larger dataset means more meaningful results. Sometimes having a massive pool of datasets actually can be more complex and confusing, and the common phrase 'analysis paralysis' can prove true. So think about that and how you might not necessarily need to focus on massive datasets.

The second misconception is the idea that 'time will tell'. However, historical data may prove to be less valuable upon further reflection. Relying on old data may not be the best predictor of the way things are going to happen in the future. Be wary of this when evaluating your data.

The third misconception is that correlation equals causation. Correlation basically means that, when one thing happens, another thing also tends to happen. So the two things are correlated together. The problem with that is that it can be difficult to discern why the second thing happened. For example, if you make a change and customers respond, are they responding because of the change you made or because of something else in the customers' psyche. When analyzing data, make sure you don't make assumptions about causation.

Ignoring the wider context is another common mistake when dealing with data. This occurs when you try to look at data through too narrow a lens. Understanding context is essential to make sure you've got the wider, holistic picture around what the data is saying.

Another misconception is that percentage and percentage points are the same. In fact, they are completely different. Be careful about that because you can experience different results if you just look at percentage against percentage points.

A final misconception is the average trap. This can occur when you have large sets of data, and people try to understand what is the average of what that data is telling you. The problem with average or median is that you don't truly understand what's happening in the various data pools so you might miss out on vital insights or you might miss out on what the data is really telling you. However, if the standard deviation from the median is very, very narrow, then the average becomes more reliable.

These are just a few key misconceptions to watch out for to make sure that you don't fall into data traps when analyzing your data. Standard deviation is a great way to overcome some of those core biases that you might come across when analyzing data.

[1.1.2.1] Now let's turn our attention to data collection and thinking specifically about some of the key considerations about collecting data. Essentially, data collection is the process of gathering and measuring information on targeted variables in an established systematic fashion, which then enables you to answer relevant questions and evaluate outcomes. By following best practices for data collection, you can ensure the trustworthiness and better use of the collected data.

[1.1.2.2] Qualitative data is effectively a depth of insights that are captured through methodologies such as interviews, for example. You should apply qualitative approaches when looking for deeper insights on already established trends or statistics, or to fill gaps in data collection that can't be filled by quantitative methods.

Common qualitative data collection techniques include:

- Interviews
- Focus groups
- Observation
- Self-study
- Ethnography
- Action research

[1.1.2.3] Quantitative analysis is looking at pure metric data to unearth insights. This can be anything from thinking about your management information that you have with the core of your organization. Sales figures or the number of complaints are good examples of quantitative measurement.

You can also look at data in terms of the customer journey from a quantitative point of view. Think about possible bottlenecks that exist within your customer journey and examine the metrics that show how many people 'drop off' at different stages to try to improve the journey. This is a very useful and unbiased way to use quantitative metrics to drive business forward.

Common quantitative data collection techniques include:

- Questionnaires
- Scales
- Physiological measurement

[1.1.2.4] We will now take a look at some of the recent or current trends around data collection. You can collect data in a number of ways, but some of the more conventional ways can be somewhat inefficient and non-responsive from a customer perspective.

Some of the more recent trends in data collection include:

- Web analytics
- Mobile market research methods
- Biometric market research techniques
- Prediction markets
- Virtual shopping
- Social listening
- The Internet of Things (IOT)

It's no longer enough to simply deliver a survey through someone's mailbox and then expect that person to complete and return it to you. You need to think carefully about what type of data you are looking for and at what point in time you capture it.

[1.1.2.5] When collecting data, you need to consider:

- Sample size
- Operational issues
- Accuracy of information
- Bias

First of all, you need to ensure that you have the appropriate *sample size*. If the sample size is wrong, the resulting data isn't reliable. For example, if the sample is too small, then some of the key insights may be missed as we have made decisions on data that is not representative of the size of the issue we are trying to analyze.

The second issue when collecting data concerns *operational issues*. Occasionally there is a difference between putting a plan down on page and implementing that plan in reality. It is important to develop ways to overcome these operational issues. You need to be clear at the beginning of the process exactly who you're looking to target.

The *accuracy* of the information is very important. If data is inaccurate it can lead to difficulties for the analysts analyzing the data and they can derive insights that are inaccurate since the starting point data wasn't accurate to begin with.

Always remember that all data collection has *bias*. Therefore, try to be aware of potential biases in your collection and overcome them as best as possible. Simply being aware of bias in data can help you to understand the data better, particularly when it does not align with predicted results.

[1.1.2.6] In general, good data is:

- Accurate
- Complete
- Consistent
- Unique
- Timely

Accuracy and completeness are essential. If there are big gaps in the data, you won't be able to draw big insights. You also need a certain consistency within your dataset. To draw insights, you want the data to be unique. In other words, you want it to be something that nobody else knows about. So you need to think about how you can capture things that you think only you have access to. And finally, as previously mentioned, good data needs to be relevant and time bound.

[1.1.2.7] In general, bad data is:

- Inaccurate
- Conflicting
- Irrelevant
- Misleading
- Outdated
- Incomplete

If the data is inaccurate, the insights based on that data will also be inaccurate. If the data is conflicting, your dataset will tell you different things, leading to confusion. If the data is irrelevant, you are simply wasting your time analyzing it. Remember to capture only what you need! Only ask for things that will be important to the questions you are seeking to answer. Outdated data can offer limited and often irrelevant details, again leading to confusion and inaccuracies. Finally, if the dataset is incomplete, it will be difficult to discern correct or reliable insights from it.

[1.1.2.8] Some common sources of bad data quality include:

- Input data
- Processing
- Data identification
- Information integration
- Usage
- Aging
- Organizational quality

Let's take a look at two common sources of bad data. How is the data being input in the first place? What are the ways in which that data could start to turn bad? At what point do you need to start to refresh that data, to ask the question again? Be careful, however. If you ask too often, you will simply annoy your customers. Conversely, if you ask too infrequently, your data may become irrelevant or too dated. Make sure that you understand the importance of getting the data right the first time.

Another consideration is process quality. You need to have a robust way of being able to process data and identify when the data is going wrong. Also consider where you store the data. You may have captured a wealth of data in one system but if it doesn't translate onto another you will encounter problems. Therefore, the usefulness of that data is compromised.

[1.1.3.1] The amount of data that's produced every day is increasing rapidly. Data processing enables us to make sense of this data. It can be defined as "the collection and manipulation of items of data on a computer to produce meaningful information". For a company or a business, this task is extremely important, as it helps to organize the information collected and makes it easier to understand.

[1.1.3.2] Data processing involves a number of micro-processes that are undertaken by analysts:

- **Data validation:** Check the accuracy and quality of source data; for example, verify metrics and remove old and incorrect data entries.
- **Data sorting:** Arrange the data in a meaningful way.
- **Data summarization:** Highlight interesting features or findings from the data.
- **Data aggregation:** Gather statistical data together and display in summary form in tables or metrics.
- **Data analysis:** Organize, evaluate, and interpret data using analytical and logical reasoning.
- **Data reporting:** Transform data and information into insights to improve business performance.
- **Data classification:** Organize data into categories for efficient use and easy retrieval.

[1.1.3.3] When it comes to processing data, analysts should follow a number of best practices:

- **Use advanced analytic techniques.** For example, analysts can leverage sophisticated statistical analysis capabilities to help identify outliers and missing values that can skew data.
- **Scrub data to build quality into the existing processes.** This involves incorporating data cleansing into your data flow and removing invalid data, to ensure your data is as robust as possible.
- **Shape data using flexible manipulation techniques.** It's important to shape your source data as you see fit. This essentially involves merging, transforming, and aggregating your data from multiple tables into one wide table.
- **Share metadata across data management and analytics domains.** By having a common metadata layer, you can repeat your data preparation processes in a consistent way. This should result in more accurate models and more transparent data.

[1.1.3.4] Let's consider some data processing challenges:

**Multiple systems:** From a technological perspective, using multiple systems can be very challenging. If you work for a big organization, you'll know that multiple systems have probably been used over time. Merging these datasets at a later point can be so difficult that it may not even be worth your time doing it! But ultimately you will want to create one version of the truth.

**Time lags:** If multiple systems are in operation and conflicting data sets exist, there may be a time lag between the data from the two systems, which analysts need to consider when processing the data.

**Inaccurate data:** Inaccurate data at the capture phase, or even at the manipulation or analysis phase, can cause problems. Analysts need to watch out for inaccurate data and use their intuition and insight to investigate further if data simply doesn't make sense.

**Data cleansing issues:** Analysts need to be able to come up with clean data, and be confident about the integrity of their data. Cleansing data to remove inaccuracies is an essential component of successful data processing. For example, updated email subscriber lists.

[1.1.3.5] Data processing also brings a number of security challenges. These can arise from four main areas.

**Physical:** There is the physical challenge of storing data, be it virtual or paper data, and where it is held.

**Personnel:** It's important to consider who has access to your data. Is it available on a need-to-know basis or does everyone in the organization have access to it? It's usually better to restrict access to those who need it.

**Procedural:** Analysts need to be aware of processing best practices and privacy policies, and think about the best ways of giving people access to data and responding to data requests.

**Technical:** Technical challenges can arise if people have difficulty using the data processing systems or software in place. In addition, transferring data from one system to another, to conduct analysis, can be very challenging from a security perspective.



## LESSON 2: ANALYTICS FUNDAMENTALS

[1.2.1.1] Web analytics involves collecting, measuring, and analyzing website data. Web Analytics tools can provide lots of useful information about the origin of website traffic, how website users navigate and interact throughout a website, what content and web pages they're most engaged with, and how they exit the site. Marketers can then use this data to optimize the performance of their channels and websites by taking data-led decisions. These are actions we take to improve performance based on our understanding of the data we've recorded; in other words, the data is leading the decision. While it's best practice to follow data-led decisions, we must always remind ourselves to make sure we test and verify the data.

[1.2.1.2] Web analytics tools provide information about the origin of website traffic, how users navigate and interact throughout a website, what content and webpages they're most engaged with, and if they take valuable actions on the site, known as conversions, these include purchases or contact requests. Using this data, marketers can optimize channel and website performance with data-led decisions.

A number of analytics tools are available on the market today. Some of them are free, and some of them require a paid subscription. Paid analytics products and free analytics products will differ in terms of support, features, and functionality.

Some examples of available analytics packages include:

- Google Analytics and Google Analytics 360, which are part of the Google Marketing Platform
- Adobe Analytics
- Woopra
- Kissmetrics
- Webtrends
- Piwik

The market leader and most commonly used analytics program is the free version of Google Analytics, or GA as it is sometimes known. The paid version of GA, called Google Analytics 360, offers some additional functionality in terms of report validity and sample sizes. But it's mainly for websites that receive more than 10 million page views per month. For the vast majority of websites, the free version of Google Analytics is perfectly fine.

Because of its functionality and widespread adoption in the market, Google Analytics is seen by many marketers as the single 'source of truth' for website traffic, engagement, and conversion data.

[1.2.1.3] Web Analytics is a valuable way to deduce the 'story' behind the data, in order to gain valuable insights and enhance business performance. Web analytics can help a digital marketer understand their customers better by providing:

- Insight into who the customers are and their interests
- Conversion challenges

- Enhanced appreciation of what consumers like or don't like
- Understanding of how to improve user experience for the consumer

One of the real values of web analytics is that it allows you to deduce the 'story' behind the data in order to gain valuable insights and enhance business performance. But how exactly does this lead to commercial returns? Web Analytics can help you understand your customers better. It tells you who they are, where they're coming from, and what their interests are. It tells you about their demographics and location. It also helps reveal any conversion challenges that might exist on your website. It helps you grasp what content and products your consumers like or don't like, and how they interact on your website. You can use all this information to improve the consumer experience on your site and to optimize the channels that consumers use to visit your website.

[1.2.2.1] When looking specifically at measuring online activity, a number of web analytics tools are available on the market today. Some of them are free, but others require a paid subscription. Paid analytics products and free analytics products will differ in terms of support, features, and functionality.

Some examples of analytics packages that are available include:

- **Google Analytics:** Google Analytics (or GA) is the most widely used free web analytics tool, with Google Ads and SEO integrations, engagement metrics, and e-commerce functionality.
- **Google Analytics 360:** Google Analytics 360 offers premium web analytics. It has similar features to the free version, but offers larger data-handling capabilities, greater reporting accuracy, and integrations with 360 Display and Video Campaign Manager.
- **Adobe Analytics:** Adobe Analytics is a premium web analytics tool that offers multi-channel integrations, tag management, voice and IOT integrations, and custom variables.
- **Woopra:** Woopra offers customer profiling, journeys, segments, and real-time analytics.
- **Kissmetrics:** Kissmetrics can be used for activity tracking, A/B testing, reporting, segmentation, and design management.
- **Webtrends:** Webtrends is used by analysts for predictive behavioral targeting, multivariate testing, visitor profiles, and segmentation.
- **HubSpot Analytics:** HubSpot Analytics can be used to analyze user journeys, conversion funnels, landing page optimization, and traffic sources.

[1.2.2.2] The market leader and most commonly used analytics program is the free version of Google Analytics. Because of its functionality and widespread adoption in the market, Google Analytics is seen by many marketers as the "single source of truth" for website traffic, engagement, and conversion data. It is effective at measuring digital marketing efforts across all channels.

Of course, websites aren't the only way people access the internet. We also have to consider the importance of mobile consumers and mobile apps. As well as collecting website data, you can collect data from mobile apps using Google Analytics SDK, or software development kit. You'll need to implement and configure additional code to view mobile app data in the Google Analytics interface. If you *are* tracking apps, you should note this requirement with your app

development team. It's not a task marketers generally do themselves, but it should be part of the brief to the app development team.

[1.2.3.1] The concept of Single Source of Truth is essential for all analytics and measurement as it is necessary to have consensus on what data is used for decision making. A single source of truth is where you determine what the core business metrics that you can measure your activity against are. It can be your CRM or your web analytics tool. It's especially important to have a single source of truth as all digital platforms measure metrics differently so you need to have a unifying set of metrics to measure different channel performance against.

Other data sources can be measured against this single source of truth data in order to optimize activities in relation to your truth. This way you are working towards an agreed data goal rather than a range of data sources that can measure actions differently.

[1.2.3.2] Using a web analytics tool such as Google Analytics as a single source of truth for conversion metrics enables you to gain a full understanding of the channel value and contribution to overall business goals. This is because all traffic, engagement, and conversion metrics can be read and analyzed side by side.

However, if conversion data is reported truthfully in another platform such as a CRM or e-commerce engine, it's important to test the variance between Google Analytics and the other single source of truth.

One way of doing this is to divide the Google Analytics data by what's been reported by your other system to see if you can confidently find a multiplier to bring that GA data in line with what's been reported in the other system. You may need to stress test the multiplier and find a reasonable compromise within an allowable variance across channels, months, or other dimensions.

[1.2.4.1] A fundamental benefit of using analytics as part of your digital marketing efforts is that it helps you make informed, data-led decisions like when to increase media spend or to stop running activity on certain online channels. This in turn should help you run effective campaigns and drive better commercial performance.

Other benefits include:

- The ability to get closer to the customer
- The ability to understand your users
- The ability to gain insights from real customer activity
- The ability to forecast using trends and patterns in customer behavior
- Accountability of media spend and resource allocation
- Stronger focus
- Prioritization of resources
- Higher conversion rates and enhanced return on investment

By measuring metrics over time in your web analytics software and noting percentage changes in metrics across areas such as channels, devices, and locations, you can demonstrate that digital activity is working when compared to historical results when the results of the comparison show a positive change.

From there, analysts can speculate and theorize about why these changes are occurring using the comparative data available to them. These hypotheses can be validated or disproven through testing, by modifying conditions such as spend level, messaging, and price.

[1.2.4.2] Google Analytics is the most commonly used free analytics tool, and it's an excellent choice for most marketing needs. Some benefits of the standard, free version of Google Analytics include the following:

- It's free.
- It's easy to implement and easy to install.
- It has a user-friendly interface.
- It offers customizable reports, dashboards, and data collection.
- It includes basic and advanced options.
- It offers seamless integration with other Google products.
- It offers post-click integration with non-Google traffic sources, including Facebook, Yahoo, and Bing.
- It provides extensive, valuable insights into web behavior.

[1.2.4.3] However, Google Analytics also has a number of limitations:

- It doesn't provide real-time data in downloadable historical reports.
- The free version is limited to 10 million hits per month.
- It offers limited customization.
- It defaults to last-click attribution.
- It's not always 100% accurate.
- It's not as comprehensive at measuring the performance of top-of-funnel campaign activities such as premium display and Facebook, where there is limited user visit and conversion data from these campaigns.

However, despite these limitations, Google Analytics is still a highly effective web analytics tool for marketers in most instances.

[1.2.5.1] The key basic terms for all analytics reporting are *dimensions* and *metrics*:

- Dimensions *describe* data. They are things such as country, channel, location, age, gender, and device used.
- Metrics *measure* data. Metrics might include visitor numbers, revenue, goal completions, and so on.

We segment metrics using dimensions – for example, visits by device type – to get a complete picture. So we can see the total visitor metrics for a website segmented by the device category dimension – that is, desktop, tablet, or mobile users. This allows us to better understand the effectiveness and scale of each device in driving visits and conversions on the website.

You should explore the metrics that are available to you, and see how they can be segmented using dimensions to get a deeper understanding of your website traffic and customers.

[1.2.5.2] Metrics can be *absolute* metrics – for example, total clicks, total revenue, and so on. Metrics can also be *relative* metrics – that is, a metric that understands one metric's relationship

to another metric (for example, conversion rate, which is the percentage of clicks that turn into conversions).

Relative metrics are created by dividing one metric by another to get, for example, the conversion rate (goals divided by clicks) or cost per click (media spend divided by clicks). Relative metrics can help with forecasting, as well as performance measurement.

In web analytics, metrics can be broadly grouped into engagement and conversion metrics.

Some common dimensions within Google Analytics include Channel Name, Month, Country, and Device. Basically, the non-numerical descriptors of what you are reporting on are considered dimensions.

[1.2.5.3] There are a number of common engagement metrics within Google Analytics:

- Absolute engagement metrics include Sessions, Users, and Page Views.
- Relative engagement metrics include Percentage New Sessions, Pages per Session, and Average Session Duration.

[1.2.5.4] Conversion metrics are an extension of standard visitor metrics. They focus on telling the story behind the valuable actions that users can take on your website. These are generally actions that impact commercial performance, such as revenue totals, conversion rates, and so on.

Google Analytics includes a number of common conversion metrics. They can all be viewed in the Traffic reports, so you can see how much value each of the different channels is driving.

Absolute conversion metrics include:

- **Goals:** These are the number of valuable actions taken on the site, including sales, leads, or downloads.
- **Transactions:** These are the number of individual sales or shopping basket items sold on an e-commerce website.
- **Revenue:** This is the monetary value of the e-commerce sales.

Relative conversion metrics include:

- **Conversion Rate:** This is the percentage of visitors that complete a goal or buy online.
- **Cost per Sale:** This is the total sales divided by total cost.

You can start to determine the value of the campaigns from your different digital marketing channels by looking at the number of goals and transactions from each channel, the conversion rates, and the cost per conversion. With these three data points, you can decide if a channel is profitable in relation to its time and cost investment.

[1.2.6.1] Remember that there are a number of data protection considerations associated with web analytics. As a result of regulations such as GDPR in the EU, tracking personally identifiable information – or PII – now requires explicit consent from the consumer. For example, you need to get consent if you are collecting cookie data from website visitors for analysis or in order to

personalize advertising experiences. Many analytics programs anonymize user names, IP addresses, and other PII, but you still need *consent* to share website visitor data anonymously with third parties, such as Google Analytics, or with ad platforms.

[1.2.6.2] In relation to GDPR, you must manage personally identifiable information correctly, and within the terms of the consent provided by your website visitors. Collecting PII as part of your analytics program, and then mismanaging that information, can risk a data breach for improper use of PII.

[1.2.6.3] To ensure this doesn't happen, your company should have a clear policy or guidelines in place detailing how the analytics data you collect is used and managed. This should include explicit opt-in visitor consent to track user visits, collect anonymous cookie data, and share anonymously with third-parties such as Google and Facebook for personalization of ads.

## LESSON 3: BIG DATA

[1.3.1.1] Big data describes the huge volume of data – both structured and unstructured – that inundates a business on a day-to-day basis. It involves datasets that are too large to store on one machine and it requires multiple computers to work together to process the volume of data. In turn, we can use the data to predict better results and forecasts than would be possible with smaller data sets.

Now, let's think about where Big Data has come from, and some of the key associated technologies around it. We can now mesh different datasets to create core insights and understanding about our customers and environment better than we've could before - it's the coming together of a number of different tools and techniques. Some of the key technologies in this area are around predictive analytics - being able to use the data to understand how things may happen in the future.

[1.3.1.2] Some technologies associated with Big Data include:

- **Predictive analytics:** using large data sets to model and outcomes based on historical data.
- **NoSQL databases:** These offer high operational speed and flexibility for data managers.
- **Search and knowledge discovery:** This involves using AI to help process and sequence data for knowledge discovery.
- **Stream analytics:** You can integrate external data sources to enable applications to integrate, update, and process additional data.
- **In-memory data fabric:** This provides a comprehensive view of business data from different sources working together through automatic orchestration.
- **Distributed file stores:** You can store files on a server for access by users across a wide network, for example.
- **Data virtualization:** This provides access to data without needing to know the technical details or where the data is stored, including how it is formatted, for example.
- **Data integration tools:** These unify data from multiple sources, giving a view of combined data.
- **Data preparation software:** This enables you to combine, structure, and organize data in preparation for data visualization, analytics, and machine learning.
- **Data quality software:** This helps you to optimize the quality of your data and deploy pre-built software to users to manage data use.

[1.3.1.3] Predictive analytics is software and/or hardware solutions that enable firms to discover, evaluate, optimize, and deploy predictive models by analyzing Big Data sources to improve business performance or mitigate risk. Effectively, this is about making core predictions based on previous behavior to evaluate what we think is the likely outcome for the future.

This is very important when thinking about buyer behavior, for example, in a marketing world. We need to be intuitive about what exactly is going to happen, and understand the variables that have changed from the past in order to make certain assumptions about what might happen in the future. It's also worthwhile using common sense and intuition when examining the data to add to the benefit of your experience of the metrics.

You can think of Big Data technologies as being like a jigsaw puzzle. It's all about meshing technologies together. There's an element of overlap, as well as linkages, because all these technologies need to come together to create a robust Big Data strategy. And if you don't have these linkages, your picture might not be complete.

[1.3.1.4] When thinking about Big Data, you can use a number of different approaches. One very useful framework involves the four Vs:

- **Volume:** the size or scale of the data
- **Variety:** different forms of data
- **Veracity:** trustworthiness of the data
- **Velocity:** frequency of incoming data

Let's start with *volume*. Big Data, by its very nature, is a larger number of data sources, and can provide greater volumes of insights on individual customers. It's the aggregating of the information that creates the richness within the dataset. In other words, it's not just a single dataset. It's about meshing a number of different data sources.

This second element is *variety*. Big Data includes a wider variety of data types and sources. For example, the data could be a rich combination of structured, semi-structured, and unstructured data.

*Veracity* refers to the need to have robust data. Essentially, you must have some form of integrity within your data. Make it trusted, make it clean, and make it de-duplicated, because any anomalies that you have within it will eventually lead to inaccuracies that you might not notice later on. Remember, robust inputs lead to robust outputs, and this applies to Big Data as well.

And, finally, *velocity*. This is the frequency of new data being entered into the set. It's always better to work with fresh data and data sets with high levels of data velocity.

[1.3.2.1] Data mining is the process of analyzing data from different perspectives and summarizing it into useful information.

It is primarily used by companies with a strong consumer focus. Think about the traditional sense of what mining is all about. It's about going deep into the center of something to get something valuable out of it. Similarly, you mine data in order to unearth core insights from it. When thinking about mining, consider these questions: What is in the data? What's the data telling me? What is the relationship between one data set and another data set?

[1.3.2.2] As previously mentioned, data mining is primarily used by companies with a strong consumer focus - retail, financial, communication, and marketing organizations, for example.

[1.3.2.3] Data mining enables companies to determine relationships among 'internal' data indicators such as price, product positioning, or staff skills, and 'external' indicators such as economic data, competition, and customer demographics.

[1.3.2.4] Data scraping is a technique in which a computer program extracts, or 'scrapes', data from human-readable output coming from another program.



[1.3.2.5] It's always worth considering: HTML content; dynamic websites; XPath and selection techniques; and regular Expressions (a language for extracting small bits of text from a larger text element), before and during data scraping.

[1.3.3.1] Data cleansing refers to identifying incomplete, incorrect, inaccurate, or irrelevant parts of a data set and then replacing, modifying, or deleting the data. In this way, outdated or incorrect information gets removed and you are left with high-quality data. For example, if you no longer sell a certain product you will want to remove its sales data from your data sets in order to focus on items that you are actively selling.

[1.3.3.2] There are a number of different methods used to cleanse data. Data cleansing may be performed interactively or on the fly by the data analyst by using data-wrangling tools and software to cleanse the data as you go, or through batch processing via scripting.

[1.3.3.3] Data cleansing helps to improve the overall quality of your data, which in turn helps you to carry out robust and effective campaigns. Remember, the validity and accuracy of your data will determine the outputs of all your analytical actions! So data cleansing is an important activity, which should never be overlooked. It plays a significant part in the role of an analyst. In fact, many analysts now spend the majority of their time cleaning data sets to ensure their accuracy.

[1.3.3.4] So how do you actually cleanse data? There are a number of steps you can follow to do this.

1: **Audit the data** using multiple methods to ensure you understand all the variables; identify good data and bad data.

2: **Consolidate the data** into a pool of good data, and a pool of bad data for removal.

3: **Establish a quality-control mechanism** where bad data is reported and removed in an ongoing way.

4: **Repeat** steps one to three.

[1.3.3.5] Data cleansing has many advantages:

**It improves the efficiency of customer acquisition.** When data is clean, customer information can be accessed more quickly and remains up to date. For example, using the wrong name in an email communication, or trying to sell someone a completely irrelevant product, can do more harm than good, and can quickly turn customers off your brand.

**It improves the decision-making process.** One of the biggest problems that organizations face is that they often don't trust the data they have. When that's the case, it becomes extremely difficult for managers to make any sort of robust decisions, as they can never be sure if the data they are basing their decision on is correct. So data cleansing becomes a critical first step before decisions can be made. It makes relevant and accurate information easily available. And it reduces the risk of misguided decisions being made on insufficient information.

**It streamlines business practices.** Once you have accurate data, you can streamline your business practices. You may be able to reduce manual processes within the organization and replace those with automated processes. For example, you could automate your email marketing communication, so thousands of emails get sent out within a few seconds. However, you can do that only when you have great, clean data.

**It increases productivity and revenue.** When you streamline business practices, this only leads to increased positivity. It also creates efficiencies within your organization that make it more cost-effective and that positively impact the firm's bottom line. When campaigns are based on clean data, you can make better decisions. As a result, campaign return on investment, or ROI, is also likely to increase.

[1.3.4.1] Data de-duplication is a process that involves optimizing data storage by eliminating *duplicate copies of data*. It's an extremely important area of data management.

Duplicate entries in data sets can artificially skew results. For example, if you have a single customer who is duplicated several times within your data set, you might overestimate the revenue value of that one individual. Alternatively, you might email the individual multiple times about a similar product, or try to cross-sell a product that the person has already bought. For best results, it's always better to work with a clean set of data that contains no duplicates.

[1.3.4.2] De-duping can occur at three stages:

- **At the client level:** This reduces the load on the central server and reduces bandwidth requirements.
- **At the server:** Data de-duplication is usually completed in the data storage layer, and it operates transparently to the clients.
- **In a block storage array of de-duping:** Data de-duping operates transparently to the clients and servers, and it is usually completed for a selected set of volumes.

[1.3.5.1] A number of different languages and programs are used by data analysts and data scientists.

**Python:** Python is a simple, general purpose programming language that's easy to work with. It allows users to write code that helps integrate data analytics and statistical code with a database. Users can use Pandas (an open-source library of data analysis tools) to help with data analytics.

**SQL:** Structured Query Language or SQL is a language that is used to manage databases. It allows users to find, update, and insert new data in large databases.

**R:** R is a free software environment and programming language that is used by statisticians and data miners. It provides a wide variety of statistical and graphical techniques that help analysts better understand their data. These include linear and nonlinear modeling, classical statistical tests, time-series analysis, classification, and clustering techniques.

**SPSS:** Statistical Package for the Social Sciences or SPSS is a statistical program that allows users to edit and analyze statistical data in a spreadsheet, and when visualized in a table or chart. It

also contains all basic statistical tests and multivariate analysis functions, such as correlation, regression, clustering, and so on.

**Hadoop:** Hadoop is a type of open-source software that allows multiple computers to be used as a network to solve problems involving massive amounts of data and computation.

[1.3.5.2] Computing languages and programs, such as those mentioned in the previous slide, can be used in Artificial Intelligence and Machine Learning. A common misconception at present is that AI and Machine Learning mean the same thing and the terms are being used interchangeably. This should not be the case; they are related but differ. We will look at both of these in more detail in the next lesson, but for now:

**Artificial Intelligence (AI)** is the ability of a computer to understand or think independently with a view to performing tasks or processes more efficiently and without the need of human intervention. AI strives towards independent decision-making by non-human systems and devices.

**Machine Learning (ML)** is a subsection of AI, using programming languages and algorithms which can adapt to new input data to drive better performance. It is the process where a computer or program can learn on its own using new data to improve its performance.

## LESSON 4: MACHINE LEARNING, AI, ANALYSIS TYPES, AND FORECASTING

[1.4.1.1] Artificial intelligence (AI) can be defined as the theory and development of computer systems that are able to perform tasks that normally require human intelligence, such as visual perception, speech recognition, decision-making, and translation between languages. Effectively, it's the cognitive processes that machines can use as a way to make key decisions based on key real-time inputs. A very simple example is a bank implementing voice and speech recognition into the internet telephone banking core services. This enables customers to, for example, access a bank balance via a voice recognition system. There are even more complicated and advanced systems being developed.

[1.4.1.2] Some catalysts for about the development of artificial intelligence include:

- Moore's Law
- Big Data
- The Internet and the cloud
- New algorithms

Moore's Law is an observation made by Gordon Moore, a co-founder of Intel, stating that the number of transistors in a dense integrated circuit doubles every two years, or the cost of manufacturing dense integrated circuits halves every two years.

The evolution of Big Data is driving artificial intelligence. With Big Data, we can put datasets together and develop predictive analysis.

The third aspect is the internet and the cloud, because they enable us to create networks of connectivity. We will talk about the Internet of Things and how that will start to feature prominently at the forefront of artificial intelligence.

Finally, creating new algorithms, or new ways of thinking about things, enables us to create better decision-making processes within our operations.

[1.4.1.3] Some common applications of AI include:

- **Finance:** fraud prevention, and avoiding accounting errors
- **Hospitals and medicine:** managing patient medical records, treatment design, and medication management
- **Customer services:** chat-bots, personalization, customer analytics, and insights
- **Transport:** driverless vehicles, traffic management, and airport facial recognition
- **Telecommunications:** network optimization, maintenance, and automation

[1.4.1.4] Some of the technologies associated with AI include:

- Computer vision
- Machine learning
- Natural language processing
- Robotics
- Facial or speech recognition

[1.4.2.1] AI has given rise to some real social concerns. These include:

- Employment issues
- Inequality
- Robotic errors
- Software security

Employment issues is one of the key concerns, particularly when machines start to replace humans' jobs across a variety of industries and fields. So how do we make sure that people have sufficient jobs? Of course, AI, or automation, doesn't render us all unemployed. Indeed, we need to think about other areas of employment that AI will struggle to fill.

The second concern is that of inequality, because there will be obvious winners and losers in this regard. We need to think about how high-intensity capital versus lower intensity capital is measured and considered to reduce inequalities. In some future predictions, the rich will get very rich and the poor will be made poorer as their jobs and modes of income are replaced by AI processes.

Nothing is perfect, so robotic errors are inevitable in an AI environment. How can we anticipate these problems and develop solutions for them?

Finally, it is important to consider software security. Recognize the issues around software security and then engage with AI to solve these problems.

[1.4.2.2] It is important to consider the economic impact of AI. These include:

- Faster decisions
- Better delegation and job allocation
- Lower costs
- Greater scale

Clearly one of the main advantages of AI is that it enables faster decision-making. This in turn enables us to respond or adapt more efficiently.

AI also results in lower costs because of reduced decision inputs and manual inputs. And it facilitates better job allocation. You can steer people away from manual tasks and have machines take care of those duties.

And finally, it creates greater scale for you. In other words, you can upscale your operation in a much more robust way, and with a lot less variability.

[1.4.2.3] The **technological singularity** is the hypothesis that the invention of AI could abruptly trigger runaway technological growth, resulting in unimaginable changes to human civilization. Fundamentally, this could possibly shift the way that humans operate, we live, the way that the economic model works, the way that we have employability, humanity, and jobs.

[1.4.2.4] Some of the benefits and limitations of AI and Machine Learning include:

### **Benefits**

Low error rates when applied correctly

Reduces human effort in repetitive tasks

24/7 application

### **Limitations**

Requires Supervision

Limited creativity and scope of improvement with experience

Still unable to replace humans

[1.4.3.1] It is important to recognize when to use AI or machine learning versus human critical thinking. Machines will offer new ways to complete tedious and repetitive tasks, but there are some tasks a machine will struggle to complete, and many tasks that a machine simply won't be able to complete as effectively as humans can. Knowing when and how to deploy machine labor and human labor is crucial to ensuring a maximum return in your strategy. Here are some examples of when to use AI and when to use human critical thinking:

### **AI**

- Repetitive Tasks
- Data sets too large for humans to process
- Situations subject to Human Bias

### **Human Critical Thinking**

- Creative Solutions
- Subjective Situations
- Empathetic emotion-based scenarios

[1.4.4.1] The Internet of Things (IOT) is a network of physical objects and devices that collect and exchange data across the internet. It is generally used to optimize the performance of the objects themselves or to provide fresh data for analysts to use. Essentially, the Internet of Things is a means to collect and process data from daily life across a variety of products from thermostats to telematics devices which report on patterns among users.

The IOT is useful for analysts as data collection is automated and specific to the collection device. This enables analysts to unearth precise insights about device use, user behavior, or other data points.

It's an ecosystem of interrelated computing devices all working on single or multiple tasks and collecting data. It provides you with unique identifiers and the ability to transfer data over a

network. Crucially, it does this without requiring human-to-human or human-to-computer interaction.

[1.4.5.1] It's important to ensure that the decisions you take can be backed up with verifiable data. This approach depends on the quality of your data, and how well it has been analyzed and interpreted. A number of different analysis methods can be used to ensure the quality of your data.

Descriptive analytics, or descriptive statistics, is a preliminary stage of data processing that describes or summarizes the basic features of a data set, and prepares it for further analysis. It uses brief descriptive coefficients to summarize a given data set and make it understandable to the reader before presenting the larger data findings. It can be either a representation of the entire population or a sample of it. Analysts use descriptive statistics to make data easier to understand and interpret.

[1.4.5.2] Descriptive statistics are broken down into:

- Measures of central tendency - how often variable gather around the mean or average
- Measures of variability - how much of variance exists in the data itself

[1.4.5.3] There are three main measures of central tendency:

- **The mean:** The arithmetic average in a data set
- **The median:** The middle value in a data set
- **The mode:** The number that is most often repeated in a data set

[1.4.5.4] Predictive analytics is another analysis type. As the name suggests, it is a branch of advanced analytics that is used to make predictions about unknown future events. It involves the use of data, statistical algorithms, and machine learning techniques to identify the likelihood of future outcomes based on historical data. For example, it can be used to predict customer responses, and to help businesses attract, retain, and grow their most profitable customers.

[1.4.5.5] Another analysis type is prescriptive analytics. It is an area of analytics that is dedicated to finding the best course of action for a given situation. It helps you draw up recommendations. It involves applying mathematical and computational sciences, and suggests ways to take advantage of the results of descriptive and predictive analytics.

[1.4.5.6] Data should be at the heart of strategic decision-making. When using data to drive your decision-making processes, follow these steps:

1. Start with your strategy and your competitors.
2. Identify your unanswered business questions.
3. Search for data you already have.
4. Work out if the costs and effort of obtaining new data are justified.
5. Collect the data.
6. Analyze the data.
7. Present and distribute the insights.
8. Incorporate the learning into the business.

[1.4.6.1] Web analytics tools can track the user or consumer journey from user origin and profile, through to on-site activity, to goals achieved and the exit point. This allows analysts to measure and understand user interactions and any valuable actions on the website using data to inspire hypotheses and insights into why these behaviors are occurring. It also allows analysts to try to recreate the conditions to drive more valuable actions or reduce the instances of poorer performing interactions.

Gaining a return from the metrics in programs such as Google Analytics is an ongoing process of monitoring, analysis, and testing enhancements from data-driven hypotheses.

[1.4.6.2] Testing allows analysts to optimize campaign performance by altering single or multiple variables at a time in a controlled way. When data insights point towards an opportunity to improve performance, analysts can create a data-driven hypothesis to test it.

Testing involves a control and a test. The control is the original set of variables; the 'test' or 'challenger' is the new set of variables. When testing two or more hypotheses, in a 50/50 test, one option will be worse than the other. Testing reduces the risk of choosing the poorer performing option for the duration of the test and once complete. The results allow for a more informed decision on whether or not to proceed with the hypothesis.

[1.4.6.3] Here are some suggested steps for testing hypotheses:

- Choose one variable to test based on your insights or data.
- Identify the goal for the test.
- Create your control (existing set of variables) and test (new set of variables).
- Set your test split (for example, 50/50).
- Set your sample size, length of test, and channel or campaign to test.
- Begin testing.

[1.4.6.4] A by-product of testing is the possibility of false positives or false negatives.

A false result is an error in the data where the result wrongly points to an outcome or condition. In testing, this is known as a type one error (false positive – where the data shows a positive result for the test when the true result was a negative), or a type two error (false negative –where the data shows a negative result when a positive result was the true outcome of the test).

[1.4.6.5] So, how can analysts test for false positives? An analyst can:

- Use common sense when interpreting results.
- Trust but verify.
- Review historical performance and look out for anomalies in the trends.

[1.4.7.1] Comparative data analysis is the item-by-item comparison of two or more comparable sets of data. It looks at the variance between the data sets over the same or a different period of time, and shows the percentage or actual change. It is an important technique for validating hypotheses or tests, and for reporting on performance.

To compare data over time:

- Choose your date range.



- Choose your data sets for each date range.
- To see the actual difference, simply subtract the data set with the older date from the data set with the newer date.
- To see the percentage difference, use this formula: new metrics minus the old metrics, divided by the old metrics multiplied by 100, equals the percentage difference.

[1.4.7.2] Analysts can use comparative data to make forecasts over time. To do this, carrying out these steps.

First, choose recent comparable data sets – for example, similar months – and work out the percentage difference.

If conditions remain the same – for example, media spend, conversion rate, and media costs – multiply this month's data by the percentage difference from the comparable period to forecast a likely outcome in simple terms.

[1.4.7.3] Analysts can also make forecasts using *historical* web metrics. For e-commerce, this involves forecasting outcomes using one or two static metrics, and one variable metric. Static variables are an assumption in the forecast.

Assume the conversion rate and average price of an item remain the same. Then you can forecast outcomes based on the number of users on the site – this is a variable metric.

Work out how many users you need to get on your site from each channel by dividing your transactions by your conversion rate to reach your objective.

Each channel has different conversion rates, so it's best practice to forecast with individual channel conversion rates, not average conversion rates. This is a common type of mistake, and is known as the Average Trap!

[1.4.7.4] You may remember the Average Trap I mentioned earlier – but what exactly is it?

Well, the Average Trap is a statistical term whereby using an average may misguide the real situation or circumstances of an event. In other words, it may result in a misleading prediction.

Ultimately, when you have large sets of data, it's an easy mistake to cling to what the average of that data is telling you. The problem with constantly going back to the average is that you might miss out on vital insights or you might miss out on what the data is really telling you.

[1.4.7.5] To ensure you don't fall into the Average Trap:

- Avoid using an average as a goal, if possible.
- If an average is a goal, check how far above or below that goal the actual performance is, relative to the goal report.

Report on the frequency of goal misses rather than on the average performance.

[1.4.8.1] Some commonly used practical algorithms include:

- Linear regression
- Logistic regression

## - Classification and regression trees

*Linear regression* is a more basic algorithm in the groups of advanced analytics. It is widely used because people can easily visualize how it is working and how the input data is related to the output data. The goal of linear regression is to identify the relationship in the form of a mathematical formula that describes the dependent variable in terms of the independent variable.

*Logistic regression* is focused on categorization of data; its goal is to categorize whether an input variable fits within a category or not. The output of logistic regression is a value between 0 and 1. The closer the result is to 1, the more of a fit the data is in that category

*Classification and regression trees* use a decision to categorize data based on questions related to one of the input variables. With each question and corresponding response, the data moves closer to being categorized.

[1.4.9.1] Pearson's Correlation Coefficient measures the strength of the relationship between two sets of numerical variables, on a scale of minus one to one, using a mathematical formula. When results are closer to minus one, there is a negative correlation; zero is a neutral correlation; and one is a positive correlation.

It is a useful method of statistically working out whether data sets are related – or correlate. This gives analysts the opportunity to decide to pursue causation tests, to see if the first set of variables *caused* the second outcome to occur. When there is a strong positive correlation, there may be causation. When there is a negative or neutral correlation, causation is less likely.

[1.4.9.2] In the majority of cases, data is understood in relation to other data sets, such as last year, other channels, other countries, and so on. The difference between the two data sets can reveal valuable insights, which in turn can lead to reporting success.

It's important that analysts fully understand the concepts of correlation (where there is a *relationship* between two events) and causation (where an event is the *result* of another event).

How closely events correlate can be determined using the Pearson Correlation Coefficient methodology. This can help show the relationship statistically in terms of positive, neutral, or negative correlation on a scale of minus one to one, using this formula which can be worked out by hand: [formula to be shown on screen]

The different parts of the formula are as follows:

**R** is the correlation coefficient number we are trying to work out. When you complete the formula, you'll find it's a number between the range of -1 to +1.

**N** is the number of data rows in your set.

**X** is your first data set.

**Y** is your second data set.

**XY** is each metric in the X data set multiplied by the corresponding metric in the Y data set.

**X squared** is the X metric multiplied by itself.

**Y squared** is the Y metric multiplied by itself.

We use the sigma symbol " $\Sigma$ " for the sum of all of the metrics in X and Y.

If you can prove a higher positive correlation between two data sets statistically, then at that point it can be worthwhile testing causation. It may not be worth testing causation for data sets with a negative correlation coefficient.

## LESSON 5: REPORTING, VISUALIZATION, AND PRESENTING DATA

[1.5.1.1] It is now time to consider how data should be reported to upper management. To begin with, there are a number of ways to report your insights and findings. Some of these include:

- Real-time access for management
- Regular updates
- Comprehensive reporting at specified points throughout the year
- Exception reporting
- So-what analysis

What are upper management really looking for in order to make decisions quickly? Well, they want access to real-time tools that you can use to help them to make decisions on a daily or even an hourly basis. You may also need to have weekly meetings or conference calls. You need to specify comprehensive reporting at specific times. Moreover, there needs to be that combination of real-time reporting versus bulk batch reporting to come up with different layers of insights. These all drive different decision-making processes.

You also need to have exception reporting because, while there will always be regular analysis, you need to think about the more exceptional incidents. For example, if anything you need to determine how an exception will be handled if it arises.

'So-what' analysis can be useful to stress test the impact and value of your recommendations and insights against common perceptions in the business. This is basically when you look at your results or your recommendations and you think "So what?". What is the implication of the analysis to the business?

[1.5.1.2] Some of the best practices for reporting data to upper management include:

- Distil the high impact findings, insights, and actions into a presentation or document.
- Keep it simple.
- Make it visually appealing.
- Have both high-level and detailed reports to hand.
- Include both quantitative and qualitative results.

[1.5.1.3] However, be mindful of what *not* to report to upper management. Here are some suggestions:

- Don't dwell on the details.
- Guide the analysis, interpret the data for them, and provide insights.
- Use words and charts to describe results for non-analysts.
- Choose the most important data for them. You don't have to give them everything!
- Remember that sometimes less is more.

[1.5.2.1] Data reporting tools allow analysts to extract and present data in different formats, such as in charts and tables, so that key information is easy to find, interpret, and action. You can use tools to visualize your data, if that's appropriate, or to create printable reports. You can create reports that are interactive, which allow users to filter and sort the data, or reports that are static. Whatever way you want to present your data; data reporting tools can help you.

Common data reporting tools include:

- **BIRT:** An open-source data-visualization tool
- **JasperReports:** An open-source Java reporting tool for creating printed documents, PDFs, spreadsheets, and so on
- **Pentaho:** A data-mining and visualization tool with dashboard capabilities.
- **KNIME:** A free and open-source data analytics, reporting, and integration platform.

[1.5.3.1] Data visualization is when you help people understand the significance of data by placing it in a visual context like a chart, a table, or a graphic. Essentially it helps to make Big Data understandable using patterns, trends, and correlations that might be undetected in text-based data.

A word cloud is a good example of data visualization. You can also use infographics to help you interpret and identify, in a visual way, key trends that you're seeing within your existing dataset. This also enables you to highlight the key insights in a visually appealing way.

[1.5.3.2] So why is data visualization important? It enables you to make big data understandable as patterns, trends, and correlations that might go undetected in text-based data can be exposed and recognized.

[1.5.3.3] Here are some of the more common tools used to visualize data:

- **Tableau:** This enables you to connect multiple data sources, and visualize data in powerful and exciting ways.
- **Microsoft Excel:** You can use spreadsheet data as the source for charts and tables.
- **Microsoft Power BI:** You can create dashboards, and interactive data visualizations with self-serve business capabilities.
- **Google Data Studio:** With this free tool, you can connect to a variety of data sources to visualize data on custom dashboards.
- **SAS Visual Analytics:** This enables you to work with self-serve analytics, visualizations, dashboards, and reports.

[1.5.3.4] Here are some advantages of data visualization:

- You can visualize relationships between operational and business activities.
- You can identify and act on emerging trends.
- It's easy to share insights.
- It's an effective way of communicating important information with non-analysts

Firstly, visualization enables you to visualize relationships between operational and business activities. It is very effective in 'joining the dots'. It also really helps the organization to move forward and understand what you're trying to achieve. Simplicity is key within these data

environments, and data visualization helps you to present insights in a simple manner. This, in turn, enables people to identify and act on emerging trends faster.

[1.5.4.1] Reporting templates are pre-built reporting structures that users simply complete to create the report. They streamline processes but are limited in the scope of reporting as they are a pre-built solution (though some have limited customization functionality).

[1.5.4.2] Reporting dashboards provide a comprehensive *snapshot* of performance. They contain curated data, charts, and tables, rather than *all* the data. And they are specifically geared towards providing visual overviews of metrics and KPIs to management, and non-analysts for the most part. They are an extremely useful way to quickly show performance and help management use data insights for better decision-making.

[1.5.4.3] There are three main type of dashboards, each with different functions, data sets, insights, and visualizations.

- **Operational:** This is the most common type, showing a snapshot of real-time performance in different channels or parts of the business - useful for management updates.
- **Analytical:** This uses historical data to try to predict future trends using comparative tables, charts, PivotTables, and data drill-downs - useful for analysts mapping trends and modelling outcomes.
- **Strategic:** This measures performance against KPIs to help decision-makers choose actions that will deliver on strategic objectives. Strategic dashboards can contain a higher amount of words than other dashboards as the audience can be senior management non-analysts, and words, rather than raw data, can be more effective in conveying the insight.