



# DATA ANALYSIS METHODS AND TECHNIQUES

---

JOKO TRILOKA, PH.D  
MTI243202

# OUTLINE

- What is data analysis and why is it important?
  - What is the difference between qualitative and quantitative data?
- 
- Data analysis techniques:
    - Regression analysis
    - Monte Carlo simulation
    - Factor analysis
    - Cohort analysis
    - Cluster analysis
    - Time series analysis
    - Sentiment analysis
  - The data analysis process
  - The best tools for data analysis

# WHAT IS DATA ANALYSIS AND WHY IS IT CRITICAL?

---

- Data analysis is, put simply, the process of discovering useful information by evaluating data. This is done through a process of inspecting, cleaning, transforming, and modeling data using analytical and statistical tools, which we will explore in detail further along in this course.
- Why is data analysis important? Analyzing data effectively helps organizations make business decisions. Nowadays, businesses constantly collect data through surveys, online tracking, online marketing analytics, collected subscription and registration data (think newsletters), and social media monitoring, among other methods.
- These data will appear as different structures, including—but not limited to—the following:



# WHAT IS DATA ANALYSIS AND WHY IS IT CRITICAL?

---

- Big data
- The concept of **big data**—data that is so large, fast, or complex, that it is difficult or impossible to process using traditional methods—gained momentum in the early 2000s. Then, Doug Laney, an industry analyst, articulated what is now known as the mainstream definition of big data as the three Vs: volume, velocity, and variety.
- **Volume:** As mentioned earlier, organizations are collecting data constantly. In the not-too-distant past, it would have been a real issue to store, but nowadays storage is cheap and takes up little space.
- **Velocity:** Received data needs to be handled promptly. With the growth of the Internet of Things, this can mean these data are coming in constantly, and at an unprecedented speed.
- **Variety:** The data being collected and stored by organizations comes in many forms, ranging from structured data—that is, more traditional, numerical data—to unstructured data—think emails, videos, audio, and so on. We'll cover structured and unstructured data a little further on.

# WHAT IS DATA ANALYSIS AND WHY IS IT CRITICAL?

---

- Metadata

This is a form of data that provides information about other data, such as an image. In everyday life you'll find this by, for example, right-clicking on a file in a folder and selecting "Get Info", which will show you information such as file size and kind, date of creation, and so on.

- Real-time data

This is data that is presented as soon as it is acquired. A good example of this is a stock market ticket, which provides information on the most-active stocks in real time.

- Machine data

This is data that is produced wholly by machines, without human instruction. An example of this could be call logs automatically generated by your smartphone.



# WHAT IS DATA ANALYSIS AND WHY IS IT CRITICAL?

---

- Quantitative and qualitative data

Quantitative data—otherwise known as structured data— may appear as a “traditional” database—that is, with rows and columns. Qualitative data—otherwise known as unstructured data—are the other types of data that don’t fit into rows and columns, which can include text, images, videos and more. We’ll discuss this further in the next section.



# WHAT IS THE DIFFERENCE BETWEEN QUANTITATIVE AND QUALITATIVE DATA?

---

- **Quantitative data is anything measurable**, comprising specific quantities and numbers. Some examples of quantitative data include sales figures, email click-through rates, number of website visitors, and percentage revenue increase. **Quantitative data** analysis techniques focus on the statistical, mathematical, or numerical analysis of (usually large) datasets. This includes the manipulation of statistical data using computational techniques and algorithms. Quantitative analysis techniques are often used to explain certain phenomena or to make predictions.



# WHAT IS THE DIFFERENCE BETWEEN QUANTITATIVE AND QUALITATIVE DATA?

---

- **Qualitative data cannot be measured objectively** and is therefore open to more subjective interpretation. Some examples of qualitative data include comments left in response to a survey question, things people have said during interviews, tweets, and other social media posts, and the text included in product reviews. With qualitative data analysis, the focus is on making sense of unstructured data (such as written text, or transcripts of spoken conversations). Often, qualitative analysis will organize the data into themes—a process which, fortunately, can be automated.



# WHAT IS THE DIFFERENCE BETWEEN QUANTITATIVE AND QUALITATIVE DATA?

---

Data analysts work with both quantitative and qualitative data, so it's important to be familiar with a variety of analysis methods.

# DATA ANALYSIS TECHNIQUES

---

## a. Regression Analysis

- Regression analysis is used to estimate the relationship between a set of variables. When conducting any **type of regression analysis**, you're looking to see if there's a correlation between a dependent variable (that's the variable or outcome you want to measure or predict) and any number of independent variables (factors that may have an impact on the dependent variable).
- Regression analysis aims to estimate how one or more variables might impact the dependent variable, to identify trends and patterns. This is especially useful for making predictions and forecasting future trends.

# DATA ANALYSIS TECHNIQUES

---

## b. Monte Carlo Simulation

- When making decisions or taking certain actions, there are a range of different possible outcomes. If you take the bus, you might get stuck in traffic. If you walk, you might get caught in the rain or bump into your chatty neighbor, potentially delaying your journey. In everyday life, we tend to briefly weigh up the pros and cons before deciding which action to take; however, when the stakes are high, it's essential to calculate, as thoroughly and accurately as possible, all the potential risks and rewards.
- Monte Carlo simulation, otherwise known as the Monte Carlo method, is a computerized technique used to generate models of possible outcomes and their probability distributions. It essentially considers a range of possible outcomes and then calculates how likely it is that each particular outcome will be realized. The Monte Carlo method is used by data analysts to conduct advanced risk analysis, allowing them to better forecast what might happen in the future and make decisions accordingly.

# DATA ANALYSIS TECHNIQUES

---

## c. Factor Analysis

- Factor analysis is a technique used to reduce a large number of variables to a smaller number of factors. It works on the basis that multiple separate, observable variables correlate with each other because they are all associated with an underlying construct. This is useful not only because it condenses large datasets into smaller, more manageable samples, but also because it helps to uncover hidden patterns. This allows you to explore concepts that cannot be easily measured or observed—such as wealth, happiness, fitness, or, for a more business-relevant example, customer loyalty and satisfaction.

# DATA ANALYSIS TECHNIQUES

---

## d. Cohort Analysis

- Cohort analysis is a data analytics technique that **groups users based on a shared characteristic**, such as the date they signed up for a service or the product they purchased. Once users are grouped into cohorts, analysts can track their behavior over time to identify trends and patterns.

# DATA ANALYSIS TECHNIQUES

---

## d. Cluster Analysis

- Cluster analysis is an exploratory technique that seeks to identify structures within a dataset. The goal of cluster analysis is to sort different data points into groups (or clusters) that are internally homogeneous and externally heterogeneous. This means that data points within a cluster are similar to each other, and dissimilar to data points in another cluster. Clustering is used to gain insight into how data is distributed in a given dataset, or as a preprocessing step for other algorithms.


# DATA ANALYSIS TECHNIQUES

---

## f. Time Series Analysis

- Time series analysis is a statistical technique used to identify trends and cycles over time. Time series data is a sequence of data points that measure the same variable at different points in time (for example, weekly sales figures or monthly email sign-ups). By looking at time-related trends, analysts can forecast how the variable of interest may fluctuate in the future.

When conducting time series analysis, the main patterns you'll be looking out for in your data are:

- **Trends:** Stable, linear increases or decreases over an extended period.
  - **Seasonality:** Predictable fluctuations in the data due to seasonal factors over a short period. For example, you might see a peak in swimwear sales in summer around the same time every year.
  - **Cyclic patterns:** Unpredictable cycles where the data fluctuates. Cyclical trends are not due to seasonality, but rather, may occur as a result of economic or industry-related conditions.
- 


# DATA ANALYSIS TECHNIQUES

---

## g. Sentiment Analysis

- Time series analysis is a statistical technique used to identify trends and cycles over time. Time series data is a sequence of data points that measure the same variable at different points in time (for example, weekly sales figures or monthly email sign-ups). By looking at time-related trends, analysts can forecast how the variable of interest may fluctuate in the future.

When conducting time series analysis, the main patterns you'll be looking out for in your data are:

- **Trends:** Stable, linear increases or decreases over an extended period.
  - **Seasonality:** Predictable fluctuations in the data due to seasonal factors over a short period. For example, you might see a peak in swimwear sales in summer around the same time every year.
  - **Cyclic patterns:** Unpredictable cycles where the data fluctuates. Cyclical trends are not due to seasonality, but rather, may occur as a result of economic or industry-related conditions.
- 

# DATA ANALYSIS TECHNIQUES

---

## g. Sentiment Analysis

There are several different types of sentiment analysis models, each with a slightly different focus. The three main types include:

- **Fine-grained sentiment analysis**

If you want to focus on opinion polarity (i.e. positive, neutral, or negative) in depth, fine-grained sentiment analysis will allow you to do so.

- **Emotion detection**

This model often uses complex machine learning algorithms to pick out various emotions from your textual data.

- **Aspect-based sentiment analysis**

This type of analysis allows you to identify what specific aspects the emotions or opinions relate to, such as a certain product feature or a new ad campaign.



# THE DATA ANALYSIS PROCESS

---

- Defining the question

The first step for any data analyst will be to define the objective of the analysis, sometimes called a 'problem statement'. Essentially, you're asking a question with regards to a business problem you're trying to solve. Once you've defined this, you'll then need to determine which data sources will help you answer this question.

- Collecting the data

Now that you've defined your objective, the next step will be to set up a strategy for collecting and aggregating the appropriate data. Will you be using quantitative (numeric) or qualitative (descriptive) data? Do these data fit into first-party, second-party, or third-party data?




# THE DATA ANALYSIS PROCESS

---

- Cleaning the data

Unfortunately, your collected data isn't automatically ready for analysis—you'll have to clean it first. As a data analyst, this phase of the process will take up the most time. During the **data cleaning** process, you will likely be:


- Removing major errors, duplicates, and outliers
  - Removing unwanted data points
  - Structuring the data—that is, fixing typos, layout issues, etc.
  - Filling in major gaps in data
- 

# THE DATA ANALYSIS PROCESS

---

- Analyzing the data

Now that we've finished cleaning the data, it's time to analyze it! Many analysis methods have already been described in this course, and it's up to you to decide which one will best suit the assigned objective. It may fall under one of the following categories:

- **Descriptive analysis**, which identifies what has already happened
  - **Diagnostic analysis**, which focuses on understanding why something has happened
  - **Predictive analysis**, which identifies future trends based on historical data
  - **Prescriptive analysis**, which allows you to make recommendations for the future
- 

# THE DATA ANALYSIS PROCESS

---

- Visualizing and sharing your findings

We're almost at the end of the road! Analyses have been made, insights have been gleaned—all that remains to be done is to share this information with others. This is usually done with a data visualization tool, such as Google Charts, or Tableau.

# THE BEST TOOLS FOR DATA ANALYSIS

---

As you can imagine, every phase of the data analysis process requires the data analyst to have a variety of tools under their belt that assist in gaining valuable insights from data. We cover these tools in greater detail in this course but, in summary, here's our best-of-the-best list:

- [Python](#)
- [R](#)
- [Jupyter Notebook](#)
- [Apache Spark](#)
- [SAS](#)
- [Microsoft Power BI](#)
- [Tableau](#)
- [KNIME](#)
- [Microsoft Excel](#)

# KEY TAKEAWAYS AND FURTHER LEARNING

---

As you can see, there are many different data analysis techniques at your disposal. In order to turn your raw data into actionable insights, it's important to consider what kind of data you have (is it qualitative or quantitative?) as well as the kinds of insights that will be useful within the given context. In this course, we've introduced seven of the most useful data analysis techniques—but there are many more out there to be discovered!

THANK YOU FOR ATTENDANCE

---

