## There are four types of data analysis that are in use across all industries:

Descriptive:

* Helps us understand what happened
* Interpretation of historical data to identify patterns
* KPI dashboard, monthly revenue report, sales lead overview, levels of customer engagement

Diagnostic:

* Helps us understand why something happened
* Builds on descriptive
* Focused on causal relationships
* Identifying correlations between variables and which are most statistically significant
* Determining factors that drive revenue, decrease turnover

Predictive:

* Attempts to answer what is likely to happen
* Uses past trends to forecast what might happen in the future
* Relies on statistical modeling. Since forecasting is an estimate, the accuracy of predictions relies on quality and detailed data
* Churn risk, sales forecasting, next best offers
  + Churn risk is the likelihood that a customer will cancel their subscription. A customer that is a high churn-risk is very likely to cancel.

Prescriptive:

* What do we need to do
* Uses optimization and simulation algorithms to
* advise on possible outcomes and prescribes that course of action
* Artificial intelligence, personalized content, dynamic pricing, self-driving cars

## Why is Data Analysis important?

Here is a list of reasons why data analysis is such a crucial part of doing business today:

* Better Customer Targeting
* You Will Know Your Target Customers Better
* Reduce Operational Costs
* Better Problem-Solving Methods:
* You Get More Accurate Data

## Data Types

What data types do we have?

* Categorical: Observation that be sorted into categories or groups. Values can be counted but not measured.
  + Nominal: Labelled categories that can be grouped but not counted
    - Marital status
    - Political party
    - Business unit
  + Ordinal: Labelled categories that can be ordered or on a scale, but the interval is not constant
    - Age group
    - Letter grade
    - Likert scale
      * A Likert scale is a rating scale used to measure opinions, attitudes, or behaviors. It consists of a statement or a question, followed by a series of five or seven answer statements. Respondents choose the option that best corresponds with how they feel about the statement or question.
* Numerical: Observations can be counted or measured. There is an order and the interval between numbers has meaning.
  + Discrete
    - Distinct values, countable
      * #of employees
      * #of orders
      * SAT scores
  + Continuous
    - Any value. This includes fractions. And the intervals have meaning
      * Value of sales
      * Market price of a product
      * Time

The course then moves into Excel to work on a datasheet called “Employee Data dataset”

Download the Employee Data dataset, a dataset about employees working at a company XYZ, and open it in Excel. Spend a few minutes exploring the different columns in the spreadsheet and their values (tip: use the filter option from the data tab). Refer to the data dictionary to have a description of the data.

## Employee data types solution

The data that are continuous don’t necessarily have decimals. We can still say that the average of **Age** is 30.5 or we can round it to 31. It all depends on the way you interpret the data and if the intervals are meaningful or not in your data analysis.

|  |  |
| --- | --- |
| **Variable** | **Data type** |
| Age | Continuous |
| Attrition | Nominal |
| BusinessTravel | Nominal |
| Department | Nominal |
| DistanceFromHome | Continuous |
| Education | Ordinal |
| EducationField | Nominal |
| EmployeeCount | Discrete |
| EmployeeID | Nominal |
| Gender | Nominal |
| JobLevel | Ordinal |
| JobRole | Nominal |
| MaritalStatus | Nominal |
| MonthlyIncome | Continuous |
| NumCompaniesWorked | Discrete |
| Over18 | Nominal |
| PercentSalaryHike | Continuous |
| StandardHours | Continuous |
| StockOptionLevel | Ordinal |
| TotalWorkingYears | Continuous |
| TrainingTimesLastYear | Continuous |
| YearsAtCompany | Continuous |
| YearsSinceLastPromotion | Continuous |
| YearsWithCurrManager | Continuous |

## The Data Analysis Process

* + 1. Ask]
    2. Collect
    3. Clean
    4. Analysis
    5. Interpret