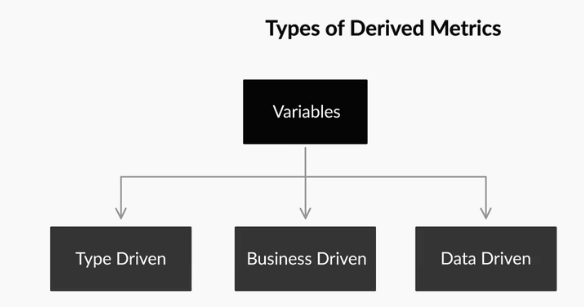
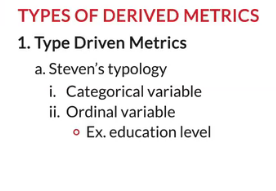
Derived Metrics

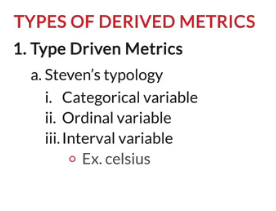
**What are Derived Metrics?**

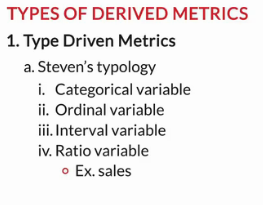
Sometimes, you would not get the most valuable insights by analysing the data available to you. You often need to create new variables using the existing ones to get meaningful insights.

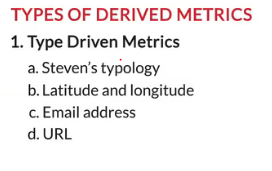
New variables could be created based on your business understanding or they can be suggested by your clients. Let’s understand how business understanding plays an important role in deriving new variables.

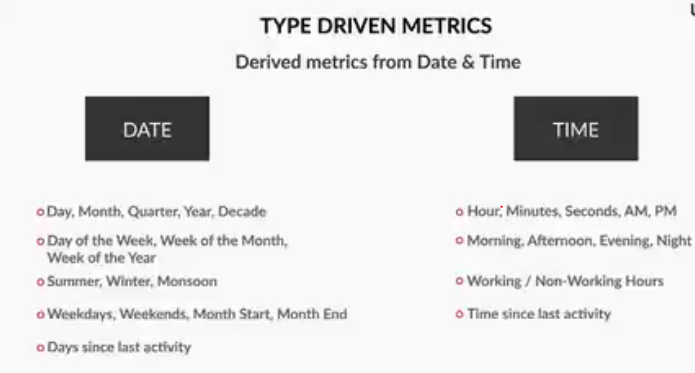












Broadly, there are three different types of derived metrics:

1.    Type-driven metrics

2.    Business-driven metrics

3.    Data-driven metrics

**Type-Driven Metrics**

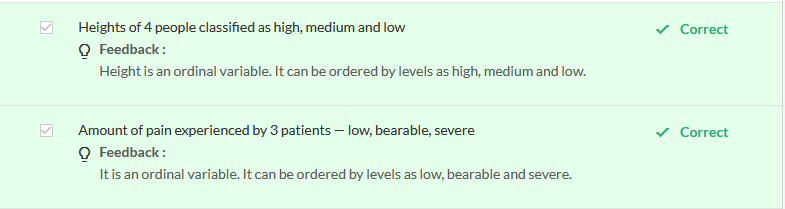
These metrics can be derived by understanding the variable’s typology. You have already learnt one simple way of classifying variables/attributes — **categorical (ordered, unordered)** and **quantitative or numeric**. Similarly, there are various other ways of classification, one of which is Steven's typology.

Steven’s typology classifies variables into four types — nominal, ordinal, interval and ratio:

* **Nominal variables**: Categorical variables, where the categories **differ only by their names**; there is **no order** among categories, e.g. colour (red, blue, green), gender (male, female), department (HR, analytics, sales)
  + These are the most basic form of categorical variables
* **Ordinal variables**: Categories follow a certain **order**, but the **mathematical difference** **between categories is not meaningful**, e.g. education level (primary school, high school, college), height (high, medium, low), performance (bad, good, excellent), etc.
  + Ordinal variables are **nominal as well**
* **Interval variables**: Categories follow a certain order, and the **mathematical difference between categories is meaningful** but division or multiplication is not, e.g. temperature in degrees celsius ( the difference between 40 and 30 degrees C is meaningful, but 30 degrees x 40 degrees is not), dates (the difference between two dates is the number of days between them, but 25th May / 5th June is meaningless), etc.
  + Interval variables are **both nominal and ordinal**
* **Ratio variables**: Apart from the mathematical difference, the ratio (division/multiplication) is possible, e.g. sales in dollars ($100 is twice $50), marks of students (50 is half of 100), etc.
  + Ratio variables are **nominal, ordinal and interval type**

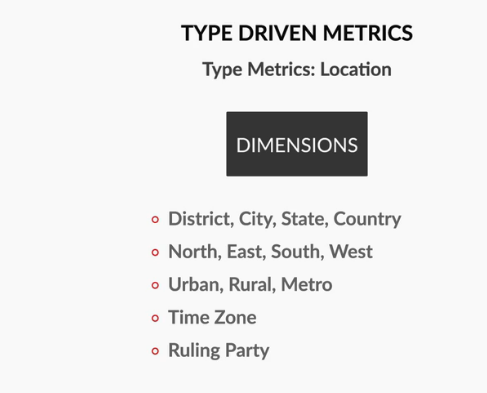
Educational degree names such as M.Tech, M.Com, MBA, MSc

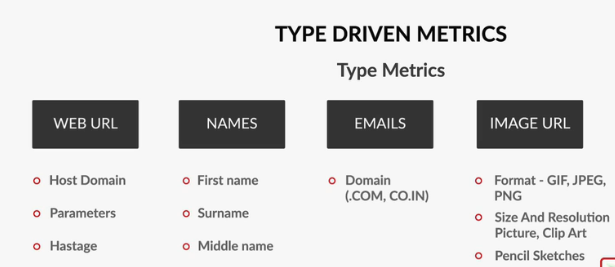
Feedback : *Nominal variables are categorical variables where the categories differ only by their names, and there is no order among the categories. E.g. education degrees such as M.tech, M.com, MBA are names of master degrees after the bachelor degree, which can be differentiated by degree names. Thus, educational degree is a nominal variable whereas “Ratings” such as “Very\_Satisfied”, “Satisfied” and “Not\_satisfied” are distinguished by labels, but these can be ordered. But “Length” is a continuous variable. It can be expressed in ratio, for which you can perform the mathematical operation on it. On the other hand, “Date of birth” can be shown in intervals. It follows a particular order, for which the measurable difference between categories is meaningful.*



Understanding types of variables enables you to derive new metrics of types different from the same column.

For example, age in years is a **ratio** attribute, but you can convert it into an **ordinal type** by binning it into categories such as children (< 13 years), teenagers (13-19 years), young adults (20-25 years), etc. This enables you to ask questions, e.g. do teenagers do X better than children, are young adults more likely to do X than the other two types, etc. Here, X is an action you are interested in measuring.





# Types of Derived Metrics: Business Driven Metrics

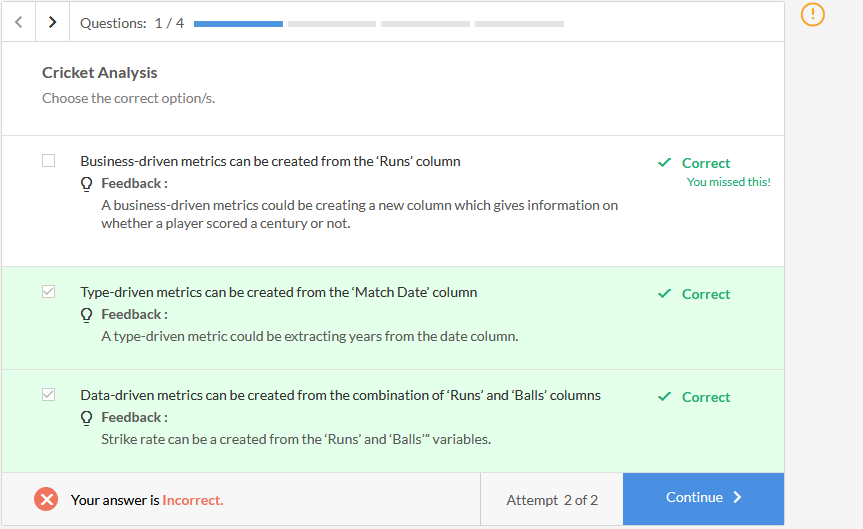
So far, you've learnt how to extract meaningful information from existing variables, e.g. extracting a "month" variable from the date variable. But what if you want to extract information that requires business expertise? For example, if you wish to know which students passed based on a list of scores in an exam, you need to know the criteria for passing the exam.



To summarise, there are **three types of derived metrics:**

1. Type-driven
2. **Business-driven**
3. Data-driven

Deriving metrics from the business perspective is not an easy task. It requires a decent domain experience. Without understanding the domain correctly, deriving insights becomes difficult and prone to errors.



Data Driven

To summarize, data-driven metrics can be created based on the variables present in the existing data set. For example, if you have two variables in your data set such as "weight" and "height" which shows a high correlation. So, instead of analyzing "**weight**" and "**height**" variables separately, you can think of deriving a new metric "Body Mass Index **(BMI)**". Once you get the BMI, you can easily categorize people based on their fitness, e.g. a BMI below 18.5 should be considered as an underweight category, while BMI above 30.0 is considered as obese, by standard norms. This is how data-driven metrics can help you discover hidden patterns out of the data.