

LESSON 2

LESSON 2.6: DATA RICH BUSINESS PROBLEMS

If we have enough data for our problem, we then find out what type of problem we are dealing with. There are two types:

- **Numeric outcome:** An outcome is a number. For Example: predicting the amount of electricity needed for each of the hours in the next day. The models used to predict these outcomes are called **Regression Models**.
- **Non-Numerical Outcome:** We try to predict the category a case falls. For example: to predict whether a customer will pay on time or predict whether a device will fail in 4000 hrs. **Classification Models** are used to predict non-numeric outcomes.

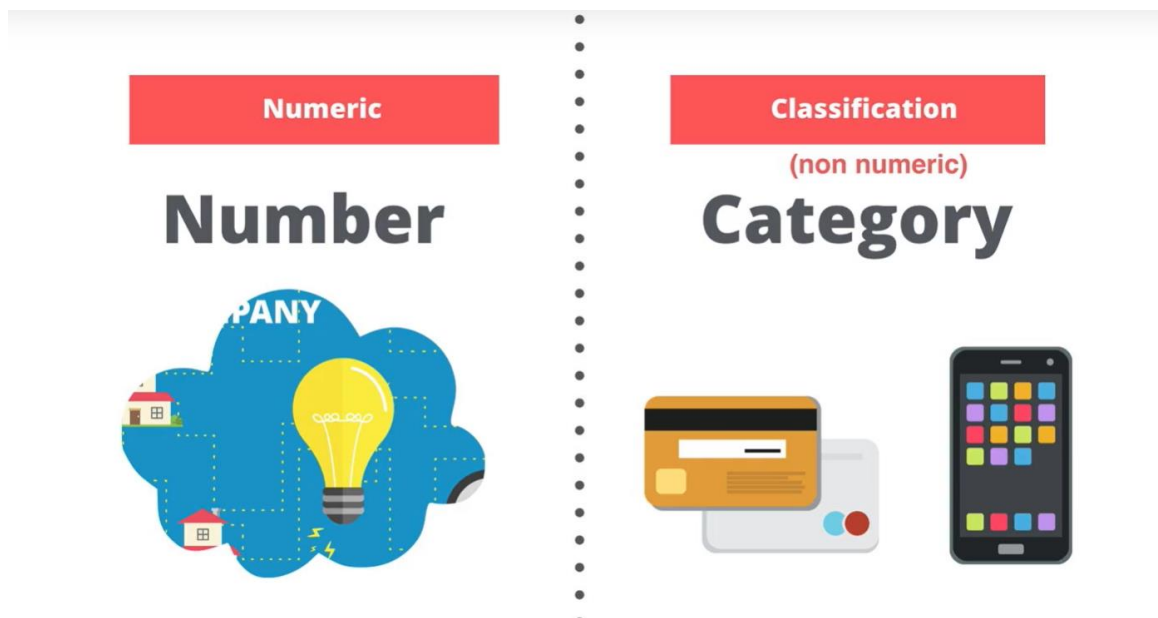


Fig 1: Types of business problems

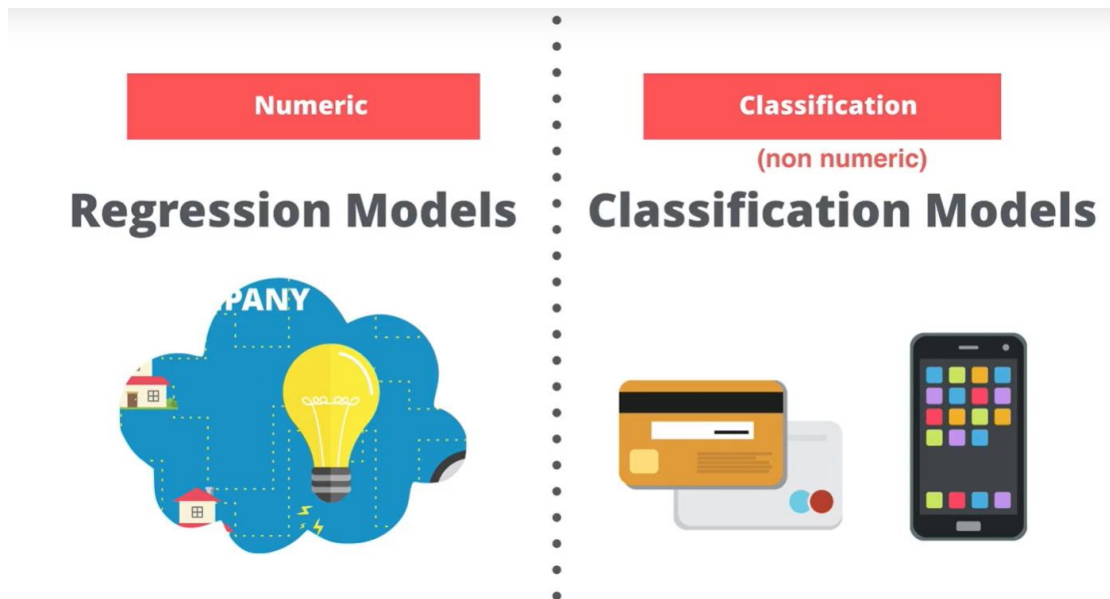


FIG 2: Models used to solve each type of business problem

2.7: NUMERIC & NON-NUMERIC OUTCOMES

1. A tricycle manufacturer wants to know how many tricycles to produce over the next six months in order to meet demand. Since the outcome is a number, we have a **Numeric problem** and we use a **Regression model** to solve this problem.

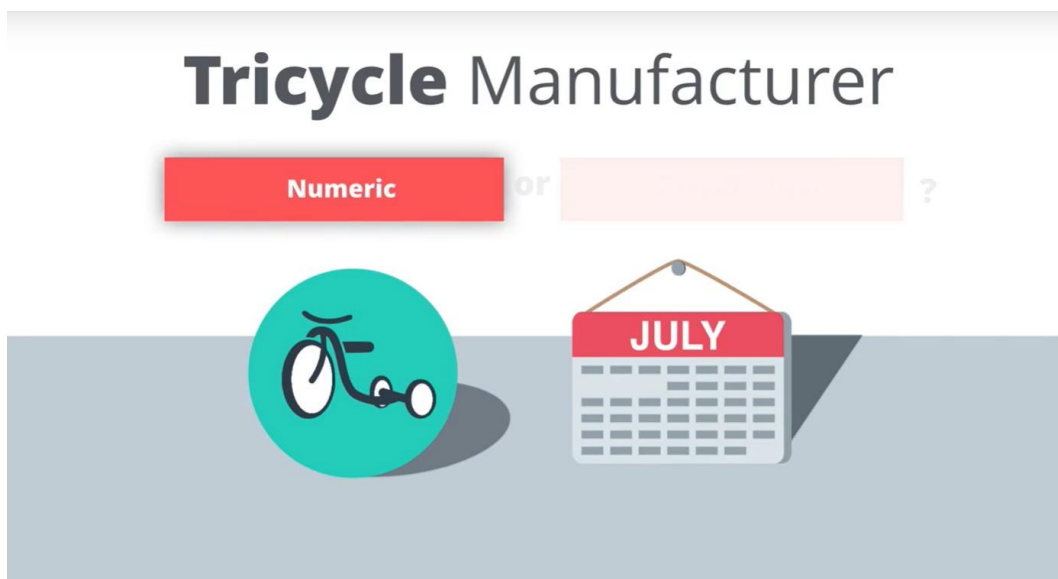


FIG 3: Tricycle problem

- Hot-and-fresh pizza, a marketing organization, wants to use sales data from existing stores and the respective demographic data around these stores to predict how many pieces of pizza they'll sell. Since the outcome they are trying to predict is also **numeric** (asking the question "how many?"), we have a **Numeric problem** and we use a **Regression model** here as well.



FIG 4: Pizza problem

- A bank wants to use historical data of their clients to predict whether a new customer will default on a loan, always pay on time or sometimes pay the loan. This problem aims to group new customers into different categories. It asks the question "How likely is this customer to default on their loan?". This type of problem is a **Classification problem** and we use a **Classification model** here.

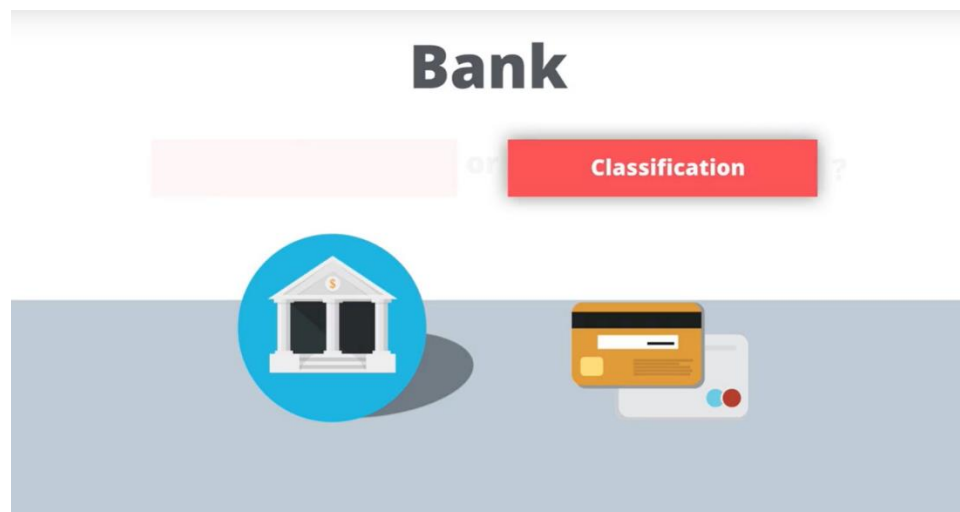


FIG 5: Bank loan problem

2.8: NUMERIC OR CLASSIFICATION QUIZ

This lesson was a quiz to test the student's understanding of Regression and Classification problems.

2.9: INTRO TO NUMERIC MODELS

Since the underlying math of a numeric model uses variables and equations, we are pretty much solving an equation for the variable that reps the outcome of our analysis. This variable is called **the Target Variable**.

Target Variables are simply variables that represent what we are trying to predict. Before selecting our predictive model, it is imperative to determine what type of target variable we are trying to predict, numeric or non-numeric. Knowing the target variable type is what helps us select the appropriate model for analysis.

TYPES OF NUMERIC VARIABLES

There are three most common types of numeric variables:

1. **Continuous:** These are variables that can take on all values in a range. An example is the height of a person, the GPA of a student, the total amount of money in your bank account. These variables can be measured down to many decimal places.
2. **Time-based:** This is usually related to forecasting. It is used when predicting what will happen over time, time series analysis. For example, forecasting the weather, predicting stock market prices or, like in our earlier example, predicting hourly electricity consumption for a given day.
3. **Count:** These variables are discrete positive integers. They are used to analyze variables that can be counted hence the name **Count numbers**. Some examples include an inventory of items in a store, the population of different countries or the number of goals scored by a soccer player in a season.

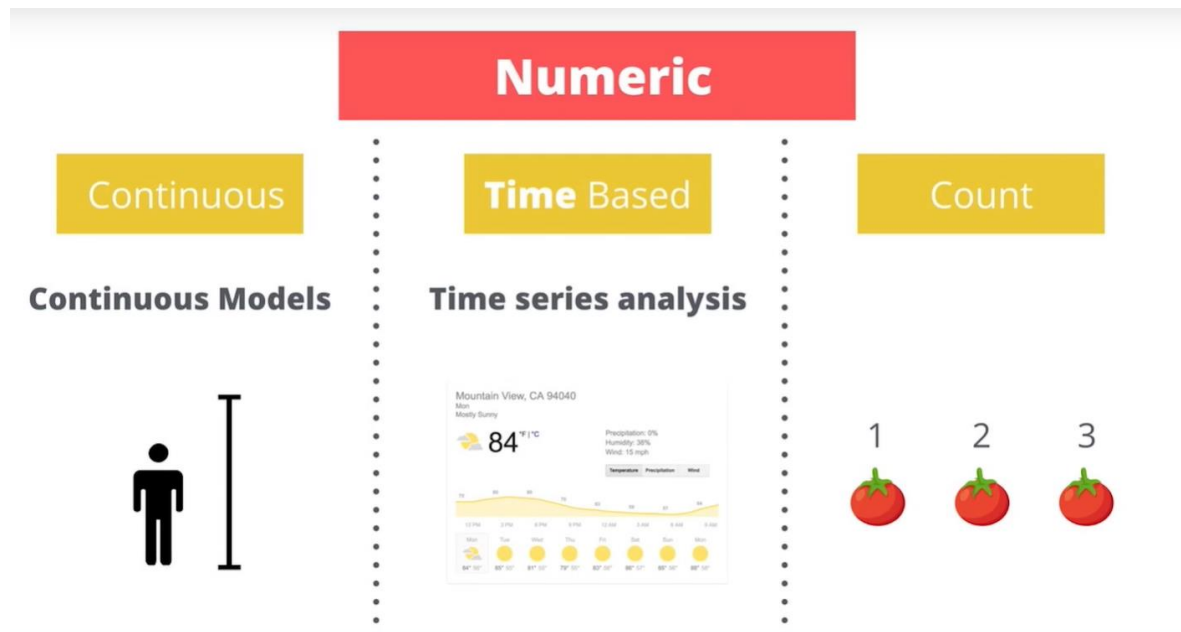


FIG 6: Types of numeric variables

2.10: INTRO TO NON-NUMERIC MODELS

Non-numeric variables take on a discrete number of possible values or categories. This is why they are called **Categorical**. For example, predicting whether or not a device will fail before 1000 hours; whether a customer will default on a loan payment or not; and predicting the classification of a flower based on its properties.

CLASSIFICATION MODELS

There are two types:

1. **Binary:** These models predict only two possible outcomes. For example: Yes or No.
2. **Non-Binary:** Unlike binary classification models, these types predict more than three outcomes. For example, predicting whether a store falls into the small, medium or large category.