# [Supplemental] Process and summarize data

As a data analyst, you're responsible for making data easier to understand. When data is first collected, it's usually in a raw, unprocessed form. Summarizing data enables analysts to extract key information from large amounts of data, and gain insights about a business, such as overall performance of a team, trends in revenue, or relationships between marketing and product sales.

In this reading, you'll learn more about how data analysts can summarize key information found in large data sets.

## Make raw data easier to understand

Raw data allows you to observe information in its original form, but by organizing data and analyzing it, you can gain deeper insights. One way that data analysts make raw data easier to understand is by calculating descriptive statistics that represent the data. Descriptive statistics are simpler, more easily interpreted values that summarize important values or trends found in large data sets. Data analysts can use descriptive statistics to monitor key performance indicators or business metrics, such as mean time spent on customer service calls, or the range of revenue per customer.

Some common descriptive statistics that data analysts use are:

- Range
- Mean
- Median
- Mode

### Range

Range is a statistic that measures the difference between the maximum and minimum values in a data set. The range of a data set is calculated by subtracting the minimum value from the maximum value. Data analysts can calculate the range to help measure how spread out the values in a data set are. For example, a data analyst exploring customer service data may want to find the range for the amount of time customer service agents spend with customers. The analyst can use the range to help measure variability in the length of customer service calls.

### Mean

A mean is a statistic that measures the central tendency or average value in a data set. Data analysts calculate the mean by finding the sum of all the values in a data set, then dividing it by the total number of data points. Means are a reliable measure of central tendency when the values in a data set are evenly spread out across the range, or there aren't any outliers. For example, a data analyst exploring how well different categories of products have sold over multiple years could find the mean revenue from each category. That way, they can quickly identify which products consistently produce the most revenue.

### Median

Another measure of central tendency in a data set is the median. The median is the middle value within a data set, and is determined by ordering the values from least to greatest to find the middle value. This value has an equal number of values that are both greater and less than it. If there's an even number of values in a data set, the two values in the middle are added together and divided by two to determine the median. Medians are useful when a data set includes outliers that are much higher or lower than most other data, because the median is less influenced by the range of the data. For example, a data analyst exploring sales data may want to determine a central tendency for prices of items sold at a store. The analyst could find the mean price, but if there are a few really high or low priced items sold, it could skew the mean value. Instead, the analyst could avoid the influence from outliers by finding the median price of items sold at the store.

### Mode

Similar to mean and median, modes also measure the central tendency of a data set. Modes are simply the most frequent or common value that occurs in a data set. Data analysts may use modes to help predict future outcomes. For example, a data analyst exploring customer feedback may want to determine the mode of customer satisfaction ratings in order to identify the most frequent satisfaction score the organization receives.

## Key takeaways

Data analysts work with raw data to make it easier to understand and analyze the data for deeper insights. Using data to calculate descriptive statistics is an important step in gaining a deeper understanding of the data. Descriptive statistics, like range, mean, median, and mode, are useful in summarizing large data sets. By effectively summarizing large data sets, data analysts can more effectively communicate important information to stakeholders to help them make data-driven business decisions.