**DATA ANALYTICS**

**Data analytics is the process of taking that data and using it to answer questions, identify trends, and extract insights that enable organizations to make better decisions.**

Data analytics involved looking at historical data to uncover trends, patterns and other descriptive information about an organization. Data sets used for this type of analysis called descriptive analytics contain static data such as customer transactions, inventory levels, or defect reports. This type of data is not usually analyzed immediately; in fact, most of it is compiled and analyzed on a monthly or quarterly basis.

The primary goal of traditional data analysis is to present information in a way that is easy for non-technical decision makers to understand. The results of descriptive analytics are highly accurate because the data usually doesn’t change during the analysis. Examples of insights uncovered through traditional analytics include past product performance, company profitability, inventory lead times, and advertising program success.

**Different Types of Analysis**

In a technical sense, data analytics can be described as the process of using data to answer questions, identify trends, and extract insights. There are multiple types of analytics that can generate information to drive innovation, improve efficiency, and mitigate risk.

There are four key types of data analytics, and each answers a different type of question:

**Descriptive analytics asks, “What happened?”**

**Predictive analytics asks, “What might happen in the future?”**

**Prescriptive analytics asks, “What should be done next?”**

**Diagnostic analytics asks, “Why did this happen?”**

Each of the above types has its own unique insights, advantages, and disadvantages. Used in combination they provide a more complete understanding of the business's needs and opportunities.

**Descriptive Analytics**

Descriptive analytics primarily uses observed data to identify key characteristics of a data set. It relies solely on historical data to provide reports on past events. This type of analysis is also used to generate ad hoc (as needed) reports that summarize large amounts of data to answer simple questions like “how much?” or “how many?” It can also be used to ask deeper questions about a specific problem. Descriptive analytics is not used to draw inferences or predictions from its findings; it is just a starting point used to inform decisions or to prepare data for further analysis.

**The descriptive analytics process is as follows:**

1. Ask a historical question that needs an answer, such as “How much of product X did we sell last year?”
2. Identify required data to answer the question
3. Collect and prepare data
4. Analyze data
5. Present results

**Examples of descriptive analytics include:**

1. Summarizing historical events such as sales, inventory, or operations data
2. Understanding engagement data such as likes and dislikes or volume of page views over time
3. Reporting general trends like revenue growth or employee injuries
4. Collating survey results

**Predictive Analytics**

Predictive analytics utilizes real-time and/or past data to make predictions based on probabilities. It can also be used to infer missing data or establish a predicted future trend. Predictive analytics uses simulation models and forecasting to suggest what could happen going forward, which can guide realistic goal setting, effective planning, management of performance expectations, and avoiding risks. This information can empower executives and managers to take a proactive and fact-based approach to strategy and decision making.

The predictive analytics process is as follows:

1. Ask a forward-thinking question, such as “Can we predict how much product X we will sell next year?”
2. Collect and prepare data
3. Develop predictive analytics models
4. Apply models to the prepared data
5. Review models and present results

**Examples of predictive analytics include:**

1. Forecasting customer behavior, purchasing patterns, and identifying sales trends
2. Predicting customer preferences and recommending products to customers based on past purchases and search history
3. Predicting the likelihood that a given customer will purchase another product or leave the store
4. Identifying possible security breaches that require further investigation
5. Predicting staffing and resourcing needs

**Prescriptive Analytics**

Prescriptive analytics builds on descriptive and predictive analysis by recommending courses of action that will reap the greatest benefit for the organization. In short, prescriptive analytics tells you what should be done in a given situation. It helps executives, managers, and employees make the best decisions based on available data.

A good example of prescriptive analytics is the field of GPS-based map and direction applications. These applications provide route options to a destination based on traffic volume, road conditions, and maximum speed. It can then prescribe the best route based on user-defined objectives such as shortest distance or quickest time.

**Diagnostic Analytics**

Diagnostic analytics enhances the descriptive analytics process by digging in deeper and attempting to discover the cause(s).

1. The diagnostic analytics process is as follows:
2. Identify anomalies (inconsistencies) in data sets
3. Collect data related to the anomalies
4. Use statistical techniques to uncover relationships and trends that could explain the anomalies
5. Present possible causes

An example of diagnostic analytics is using subscription cancellations, correlated with customer comments and ratings, to determine the most common reasons why users cancel subscriptions. Another example would be determining whether there is a correlation between the demographics of consumers and their purchasing patterns at specific times of year.

**The Data Analysis Process**



* **Asking the Question** - The analytics process always starts with a question to be answered. Some questions are simple, such as “which bicycle color is most popular with our younger buyers?” Others are very complex, such as “why are certain types of cancer cells exhibiting resistance to radiation treatments?”
* **Getting the data** - This step involves the process of locating and obtaining data that is relevant to the question, and then determining if there is enough data to complete the analysis.
* **Investigating the data** - Data comes in many forms and from many different sources. This step involves determining if the data is complete and contains the relevant information for the analysis.
* **Preparing the data** - This step can involve many tasks to transform the data into a format appropriate for the tools that will be used to analyze and present the data. This process is sometimes referred to as “cleaning” the data, because there may be blank records or obvious errors in the data set.
* **Analyzing the data** - Analysis is the process of identifying patterns, correlations, and relationships contained within a data set or sets to draw inferences and conclusions. Often, analysis relies on statistical techniques and software tools such as spreadsheets and visualization applications.
* **Presenting the results** - This is usually the last step for data analysts. It is the process of communicating the results to decision-makers. This can be done in the form of a report, graphical representations, or a combination of both. Sometimes the data analyst is also asked to recommend actions based on results.

**Use simple functions for data analysis**

**Observations, Variables, and Values**

When performing any kind of data experiment or analysis, it is critical to define the key characteristics that need to be measured or observed. These characteristics to be studied are called variables. A variable in this context is anything that varies from one instance to another, that can be measured, and whose value can be manipulated or controlled in theoretical scenarios.

The recordings of the values, patterns, and occurrences for a set of variables are observations. The value or set of values for a specific observation is called a data point. Each observation can be thought of and represented as a record in a database or a row in an Excel spreadsheet. The collection of observations makes up the dataset for your analysis.

Observations usually have a purpose, and the variables included will depend on their relevance to that purpose. For example, if you have lost your pet and have asked other people to help you search for it, only a small set of variables—the dog’s characteristics—are relevant to their observations. These characteristics might be:

What type of animal is your pet?

It is a dog.

What type of dog?

It is a Schnauzer.

What color is your Schnauzer?

It is gray.

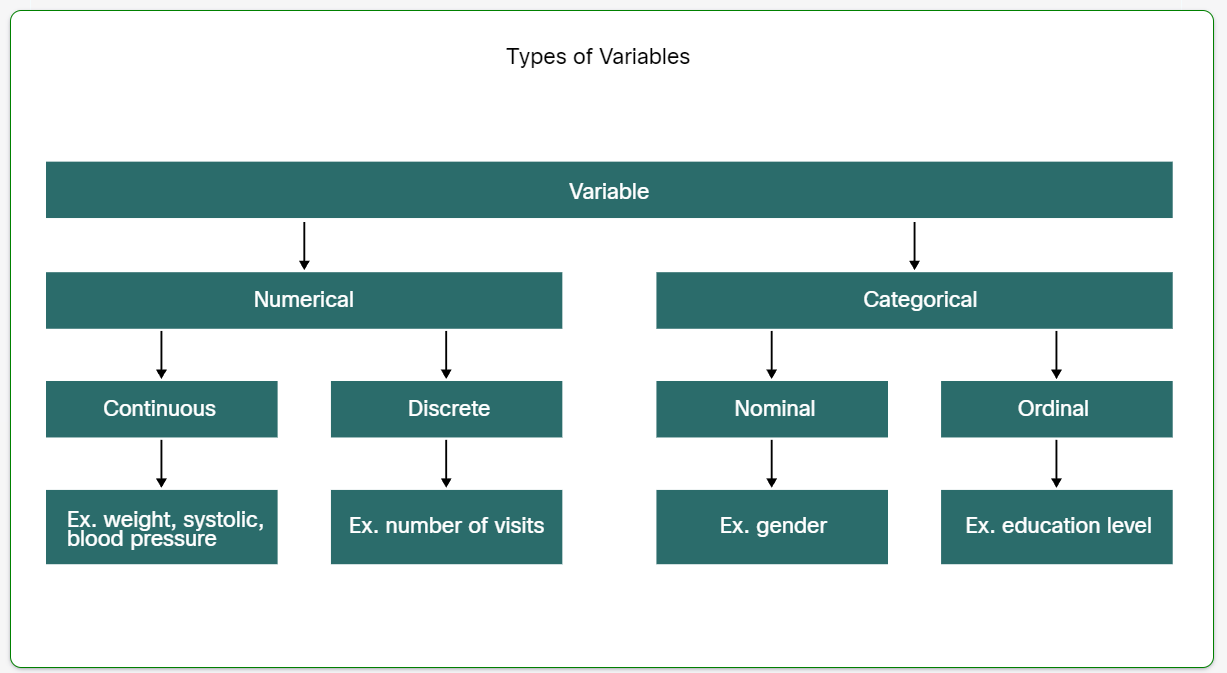
What size is the Schnauzer?

It is a medium sized Schnauzer.

How much does the Schnauzer weigh?

It weighs 15 kg.

**Types of Variables**

****

Categorical variables indicate membership in a particular group and have a discrete or specific qualitative value.

They are further classified into two types:

* **Nominal** - These are variables that consist of two or more discrete categories whose value is assigned based on the identity of the object. Examples are gender, eye color or type of animal.
* **Ordinal** - These are variables that consist of two or more categories in which order matters in the value. Examples are student class rank (1st, 2nd, 3rd) or satisfaction survey scales (dissatisfied, neutral, satisfied).

Numerical variables are quantitative values:

* **Continuous** - These are variables that are quantitative and can be measured along a continuum or range of values. There are two types of continuous variables: Interval variables can have any value within the range of values, and examples are temperature or time; Ratio variables are special interval variables where a value of zero (0) can mean that there is none of that variable and examples are income or sales volume.
* **Discrete** - These types of continuous variables are quantitative but have a specific value from a finite set of values. Examples include the number of sensors activated in a network, or the number of cars in a lot.

**Sources of Data**

**Selecting Relevant Data**

Selecting relevant data for your analysis includes determining the type(s) of data that you need and finding a source for the data. When selecting data for a project, it is important to focus on finding data that may provide insights into your original business question. For example, if you are seeking to understand demographic characteristics of people who bought Product X in the past year, you should only be using data that is directly related to Product X. This process is crucial to ensuring the validity and reliability of your analysis. Sometimes the data you need to answer your questions isn’t readily available. It may be necessary to establish new procedures to collect the data required for your analysis. Other times, it may involve combining data from multiple sources into a format that can be analyzed.

**Some questions that you should ask yourself when selecting a data source:**

1. What data points are necessary to inform your analysis?
2. Do I already have access to this data, or must I find a dataset from another source?
3. Where are reliable and verifiable sources of this data?
4. How often is the relevant data collected and updated?
5. How is the data licensed for use, and is there a cost?
6. Is the data in a format that I can use, or convert to use, with my tools?

**Static and Streaming Data**

There are two types of data that analysts work with: static data and streaming data. Data that is received and stored prior to performing analysis on the data is considered static data. When each event is processed and analyzed as it is received and subsequent results are used or stored, the data is referred to as streaming data.

**Data in Structured Files**

Structured data refers to data that is entered and maintained in defined fields within a file or record. Structured data is easily entered, classified, queried, and analyzed by a computer. This includes data found in relational databases and spreadsheets. For example, when you submit your name, address, and billing information to a website, you are creating structured data. The database may force you to enter it in a certain format for a computer to interpret it easily.

**Some of the characteristics of structured data include:**

* There is a well-defined and organized structure.
* It can be stored in tables, usually within vertical columns and horizontal rows.
* The content and format of the data is documented.
* It is organized into files, records, and fields.
* It can be searched, sorted, and queried.
* Input controls can reduce the possibility of invalid data.

**Unstructured Data**

Unstructured data lacks the organization found in structured data. Unstructured data is raw data, not organized in a predefined way. It does not possess a fixed schema that identifies the type of data or its format. This type of data lacks a set way of entering or grouping the data, and then analyzing the data.

Examples of unstructured data include the content of photos, audio, video, web pages, blogs, books, journals, white papers, PowerPoint presentations, articles, email, wikis, word processing documents, and text in general. Even the PDF version of this chapter is unstructured. The text is searchable, but it is not organized in a predefined format. Unstructured data can even be a traffic camera feed that is continuously sending images for processing.

**Both structured and unstructured data are valuable to individuals, organizations, industries, and governments. It is important for organizations to take all forms of data and determine ways to format that data so it can be managed and analyzed.**

**Data Preparation**

**ETL and ELT Processes**

ETL and ELT are two versions of the same process for moving data through a pipeline. They contain the same steps but in different orders for different use cases.

Extract, Transform and Load (ETL) is a process for collecting data from this variety of sources, transforming the data, and then loading the data into a database. One company’s data might be found in Word documents, spreadsheets, plain text, PowerPoints, emails and PDF files. Another company’s data may be housed in relational databases. This data can be stored in a variety of different formats, making it difficult to combine and analyze, so the transformation happens before loading.

In an Extract, Load, Transform (ELT) process, the load and transform steps are reversed. ELT enables raw data to skip the transformation step and go straight to storage in an unstructured form. Transformation then occurs on the stored data as it is used. The ELT process is used primarily for large amounts of unstructured data.

**Let’s take a closer look at the three steps in the ETL process.**

**Step 1 : Extract**

In this step, data is located and gathered from various sources in order to be converted into a single format for analysis. The data may be extracted from a relational database, NoSQL, flat files, XML files, or other formats.

**Step 2 : Transform**

Data usually must be transformed before it can be loaded into a data warehouse for analysis. The transform step uses rules to transform the source data to the type of data needed for the target database. This includes converting any measured data to the same dimension (e.g., Imperial to Metric). The transformation step also requires several additional tasks. Some of these tasks are joining data from several sources, aggregating, sorting, determining new values that are calculated from aggregated data, and then applying validation rules.

Data (possibly including some empty or error data) may go through another part of the transform step known as ‘cleaning’ or ‘scrubbing’ data, and validation lets you know whether the data needs cleaning. Some examples of data cleaning are removing blank records and standardizing formats such as date, time, and location. The cleaning part of the transform step further ensures the consistency of the source data.

**Step 3 : Load**

The transformed data is then loaded into the database for querying. The actual load process varies widely, depending on the types of source data, the type of target database, and the type of querying that is to be done. Some organizations may also overwrite existing data with newer cumulative data. During the load step, rules that have been defined in the database schema are applied. These rules check for necessary characteristics like uniqueness and consistency of data or that mandatory fields aren’t empty. These rules help to ensure that the loading and any subsequent querying of the data is successful.

**Formulas and Functions**

Excel supports a wide range of formulas and function capabilities—the only difference between these is that functions use keywords and are predefined by Excel, while users can build custom formulas themselves using operators. There are strict rules to follow when creating formulas, so it is a good idea to plan out how you want the calculations to work before you start entering them into your spreadsheet.

First, the basics: all Excel formulas start with an equal sign **(=)**. The equals sign indicates that the data in the cell is not text or a number, but a calculation. This ensures that your formula will not be displayed in the cell instead of the result of the calculation. For example, **=1+2** entered into a cell (with the equals sign as the first character) will display the value **3**, not the text *=1+2*.

**Relative and Absolute References**

Calculations can be performed on either a constant, like the number 5, or the data contained at a specific location (usually a cell or range of cells). For example, the formula **=L1/L7** divides the value in cell **L1** by the value in cell **L7**.

Within a formula or function, when there are cell or range references, they can be either relative or absolute. A reference is relative when its value depends on the location of the reference itself, while an absolute reference refers to the same cell or range no matter where the reference appears. For example, when you did the Manipulate Data lab, you copied and pasted formulas from one row into the rows below it. When you did that, because the formulas had relative references, they used the values from the lower rows to make the calculations for those rows. This is the default behavior for formulas and functions.

If you want a formula to refer to the same cell or range of cells no matter where the formula is pasted, you can make the reference absolute by adding a dollar sign ($) before the column or row indicator for that reference. Let’s use the example referenced above, **=L1/L7**. If you wanted to use this formula in multiple places in your spreadsheet and always have it use the values in L1 and L7, rather than the values that correspond to the formula’s new location, you would write it as =$L$1/$L$7.

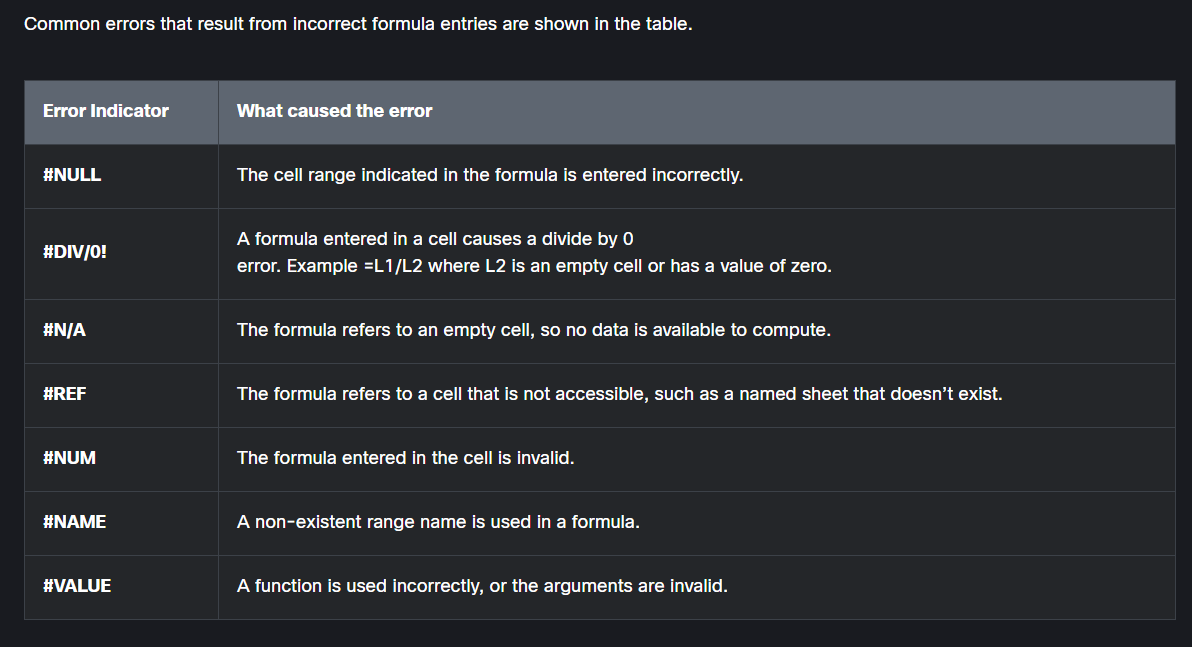
Both the column indication and the row indication can have this attribute, independently of each other, so you can also add the dollar sign to only the column reference or only the row reference.

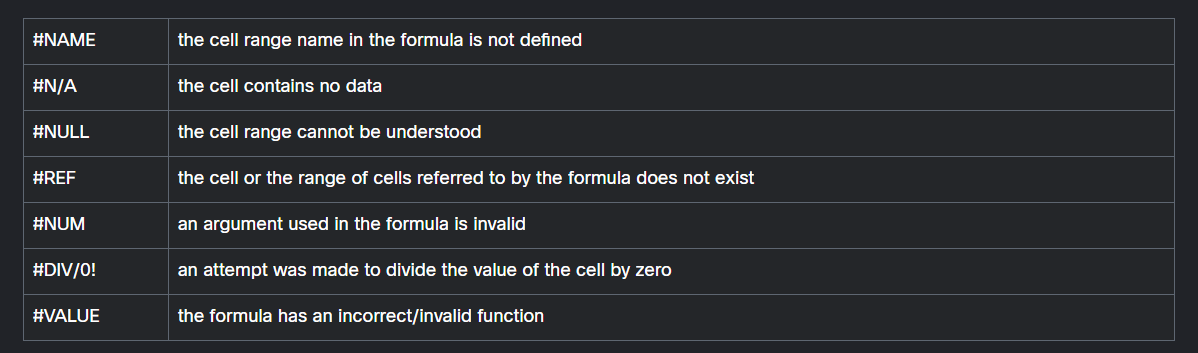
**Functions**

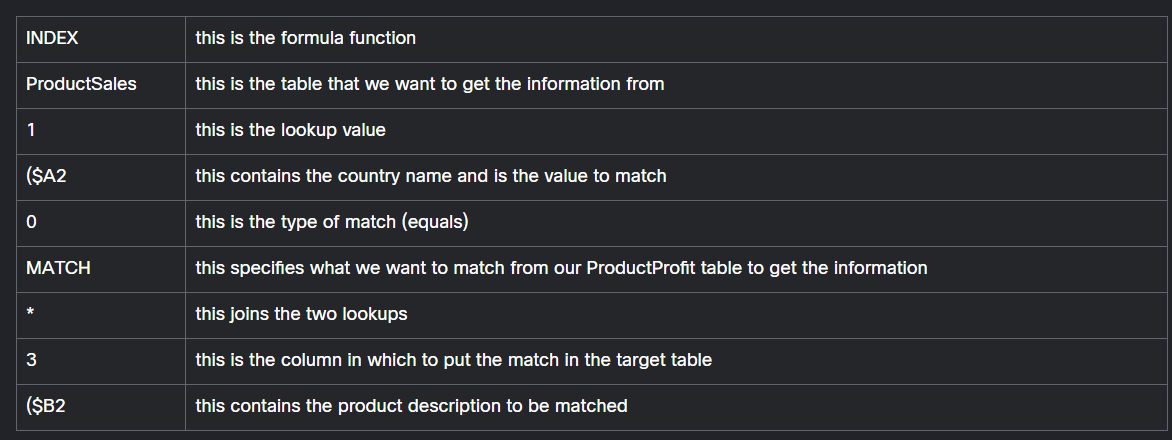
Built-in functions are formulas that start with a keyword that identifies a specific function to be performed. Most function keywords describe the calculation that the function will perform. Correct syntax varies by function, but usually has a similar pattern. For example, the syntax for the AVERAGE function accepts a range or list of cells to calculate the average of the values within the range, such as **=AVERAGE(L2:L37)** or **=AVERAGE(L2, L5, L6)**.

**Note:** You must make sure to close all parenthesis and brackets, or you will receive an error message when you press enter to complete the entry. For long functions this can be difficult.

When you start a cell entry with an equal sign, as you type the function name, a menu of functions matching your entry will usually appear and show the short description of each function listed. If this doesn’t happen automatically, you can also right-click on a cell and enter the equals sign in the text field that says “Search the menus” to show this menu. Selecting the function that you want places the function name and an open parenthesis on the formula bar and in the cell.







**Pivot Tables**

Pivot tables are a function in Excel that is invaluable to data analysts. Pivot tables provide a way to automatically summarize, analyze, explore, and present data. Pivot charts enable you to add visualizations to the data in a pivot table. Using these built-in tools you can identify trends, make comparisons between data items, and create charts in different styles to visualize your data. You’ll be creating a Pivot Table in the upcoming lab, but for reference, here are the steps to create a pivot table in Excel:

1. Start with a worksheet that is organized in columns and rows, with column headers.
2. Select the table or range of cells that you want to be included in your Pivot table. Be sure to include the column headers in the range.
3. Choose Insert from the menu bar. In the Tables section, you can view various formats that your data can take by selecting the Recommended Pivot tables choice. If you don’t see a recommended table that meets your requirements, select Pivot table => From Table or Range.
4. The Create Pivot table dialog box opens. Your selected range should show in the Table/Range box, and New Worksheet should be selected. When creating your first Pivot tables, it is best to have them placed on a new worksheet tab.
5. A blank Pivot table worksheet is created and the Pivot table Fields list containing your column headings appears. Click the fields that you want to add.
6. The Pivot table wizard will place your fields in one of the boxes at the bottom, Filters, Columns, Rows, and Values. You can drag and drop your fields into different categories to change the way that your Pivot table summarizes the data.
7. The Pivot table will appear with your selections. Pivot tables support most of the functions that you can do on a normal spreadsheet, such as sorting, filtering, and cell formatting.

Updating data in the original spreadsheet does not update the pivot table automatically; you must refresh your Pivot table to have it reflect new data.

**Descriptive Statistics**

After the problem statement (also known as the question to be asked) and population is determined, some form of statistical analysis is needed. There are two key branches of statistics that we will discuss in this course:

* Descriptive Statistics
* Inferential Statistics

Descriptive statistics are used to describe or summarize the values and observations of a data set. For example, a fitness tracker logged a person’s daily steps and heart rate for a 10-day period. If the person met their fitness goals in 6 out of the 10 days, then they were successful 60% of the time. Over that 10-day period, you could observe that the person’s heart rate was a maximum of 140 beats per minute (bpm), but an average of 72 bpm. These observations would be descriptive statistics that could be used to describe and simplify the data set.

Basic descriptive statistics might include the total number of data points in a data set, the range of values that exist for those numeric data points, or the number of times a given value appears in a data set. Descriptive statistics may also answer questions about the occurrence of trends.

The answers to these questions can be provided in numerical or graphical formats. Results of descriptive statistics are often represented in pie charts, bar charts or histograms. One important point to note is that while descriptive statistics describe the current or historical state of the observed population, it does not allow for:

* comparison of groups
* conclusions to be drawn
* predictions to be made about data sets that are not in the population

In the fitness tracker example, we cannot infer that the person has poor health because they were only successful in meeting their goal 60% of the time. We also cannot use the data set for this one person to predict the fitness performance for others with similar characteristics. This is where inferential statistics becomes important.

**Inferential Statistics**

Descriptive statistics allows you to summarize findings based on data that you already have recorded or observed about a population. However, there are situations in which gathering data for a very large population may not always be practical or even possible. It is possible, however, to study a smaller representative sample of a population and use inferential statistics to test hypotheses and draw conclusions about the larger population.

Inferential statistics is the process of collecting, analyzing and interpreting the data gathered from a sample to generalize or predict something about a population. When a representative sample is used, methodological concerns may arise and must be addressed, such as whether the groups chosen for the study or the environment in which a study is carried out accurately reflects characteristics of the larger group. Typically, these types of analyses will include different sampling techniques to reduce error and increase confidence in the generalized findings. The type of sampling technique used will depend on the type of data.

**Statistics and Big Data**

Different statistical approaches are used in big data analytics. As we know, descriptive statistics describe a sample. This is useful for understanding the sample data and for determining the quality of the data. Problems can occur when dealing with large amounts of data that come from multiple sources. Data points can be corrupted, incomplete, or missing entirely. Descriptive statistics can help determine how much of the data in the sample is good for the analysis and identify criteria for removing data that is inappropriate or problematic. Graphs of descriptive statistics are a helpful way to make quick judgements about the quality of a sample.

For example, in a sample of tweets selected for analysis, some contain only text characters, while others contain both characters and images. The type of analysis or question to be answered with analysis will determine whether tweets that contain images or tweets with no images should be analyzed. This will identify tweets that are invalid based on a very simple criterion, because images contain information that must be considered in the analysis if the tweets using images are included in the sample.

A number of inferential analyses are very commonly used in big data analytics:

* Cluster analysis - Used to find groups of observations that are similar to each other
* Association analysis - Used to find co-occurrences of values for different variables
* Regression analysis - Used to quantify the relationship, if any, between the variations of one or more variables

**Visualization**

**Common Types of Data Visualizations**

There are many types of data visualizations. Determining the best option usually depends on the answers to the following questions, among others:

How many variables are you going to show?

How many data points are in each variable?

Is your data over time or are you comparing data points at a single point in time?

Select each chart type to review some uses and best practices.

* Line Charts
* Column Charts
* Bar Charts
* Pie Charts
* Scatter plot

**Line Chart**

Line charts are one of the most commonly used types of comparison charts. Use line charts when you have a continuous set of data, the number of data points is high, and/or you would like to show a trend in the data over time. Some examples include:

* Quarterly sales for the past five years
* Number of customers per week in the first year of a new retail shop
* Change in a stock’s price on one day, from opening to closing bell

Some best practices for line charts include:

* Label the axes.
* Plot time on the x-axis (horizontal) and the data values on the y-axis (vertical). Use a solid line (rather than a broken line) to emphasize continuity of the data.
* Keep the number of data sets to a minimum. There should be a very good reason for plotting more than four lines. If needed, add a legend to help the audience understand what they are viewing.
* Remove or minimize gridlines to reduce distraction. Consider using no gridlines except to emphasize certain values or time periods.
* Modify the y-axis starting point to obtain something close to a 45-degree slope in one or more of the lines. This ensures you emphasize the change in the data without introducing distortions that dramatize the visualization.

**Column Chart**

Column charts are positioned vertically. They are probably the most common chart type used to display the values of a specific variable across similar categories. Some examples include:

* Populations of the BRICS nations (Brazil, Russia, India, China, and South Africa)
* Last year’s sales for the top four car companies
* Average student test scores for six math classes

Some best practices for column charts include:

* Label the axes.
* If changes over time are being shown, time should be plotted on the x-axis.
* If time is not part of the data, consider ordering the data so that column heights ascend or descend.
* Fill the columns with a solid color. To highlight one column, consider using an accent color and make all the other columns the same color.
* Column charts are best when there are no more than seven categories on the horizontal axis. This will help the viewer clearly see the value for each column.
* Start the value of the y-axis at zero to accurately reflect the full value of each column.
* The spacing between columns should ideally be roughly half the width of a column.

**Bar Chart**

Bar charts are similar to column charts except they are positioned horizontally and hence used slightly differently (for example, they do not usually show changes over time). Longer bars indicate larger values. They are best used when the names for each data point is long, because there is space to write the information. Some examples include:

* Gross domestic product (GDP) of the 25 highest-producing nations in a given year
* Number of cars sold by each sales representative in a group
* Exam scores for each student in a math class

Some best practices for bar charts include:

* Label the axes.
* Consider ordering the bars so that the lengths go from longest to shortest. The meaning of the data shown will most likely determine whether the longest bar should be on the bottom or the top for greatest impact or easiest understanding.
* Fill the bars with a solid color. To highlight one bar, consider using an accent color and make all the other bars the same color.
* Start the value of the x-axis at zero to accurately reflect the full value of each bar.
* The spacing between bars should ideally be roughly half the width of a bar.

**Pie Chart**

Pie charts are used to show the composition of a total. Segments of different sizes visually represent percentages of that total. The sum of the segments must equal 100%.

Some examples include:

* Annual expenses for a corporation (e.g., rent, administrative, utilities, production)
* A country’s energy sources (e.g., oil, coal, gas, solar, wind)
* Survey results for a group’s favorite type of movie (e.g., action, romance, comedy, drama, science fiction)

Some best practices for pie charts include:

* Limit the number of categories so that the viewer can easily differentiate between segments and their meaning in relation to each other. After ten or more segments, the slices begin to lose meaning and impact.
* If necessary, consolidate smaller segments into one segment with a label such as “Other” or “Miscellaneous”.
* Use a different color or gray scale for each segment.
* Order the segments clockwise according to size.
* Make sure the value of all segments equals 100%.

**Scatter Plot**

Scatter plots are very popular for visualizing correlations, or to show the distribution of many data points. Scatter plots are also useful for demonstrating clustering or identifying outliers in the data.

Some examples include:

* Comparing life expectancy to GDP for each country in a group
* Comparing the daily sales of ice cream at a given location to the average outside temperature
* Comparing the weight to the height of each person in a group

Some best practices for scatter plots include:

* Label the axes.
* Make sure the data set is large enough to provide visualization for clustering or outliers.
* Start the value of the y-axis at zero to accurately reflect the full values of the data. The value of the x-axis will depend on the data. For example, age ranges of ice cream customers might be labeled on the x-axis, and there would be no need to start at zero years of age.
* If scatter plot shows a correlation between values on the x- and y-axes, consider adding a trend line.
* Do not include more than two trend lines.

**VLOOKUP**

VLOOKUP is a very powerful data analysis tool within Excel and is great when you need to find information in a large spreadsheet or if you are consistently looking for the same type of information.

VLOOKUP is an abbreviation of “vertical lookup,” and it’s a function that searches a (vertical) column in a table for a specified value. This means that the data must be organized in a table where each row has different but related forms of data in each column. If an approximate match is specified in the formula, the first column (the lookup column) must be sorted in numeric or alphabetic order.

A VLOOKUP function consists of 4 key pieces of information:

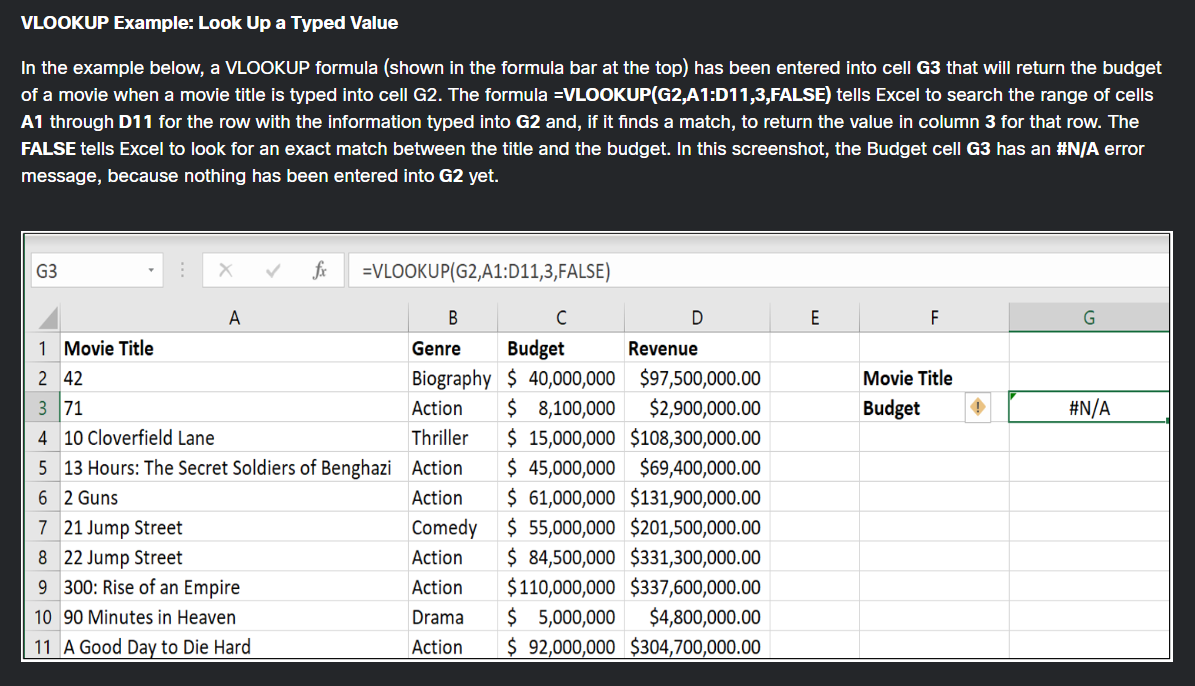
1. The value to search for
2. The range to search in
3. The column in the range that contains the value you want the function to return
4. An indication of whether the function should return an approximate match (TRUE, in the function) or only an exact match (FALSE) of the return value. The default for VLOOKUP is an approximate match if FALSE is not specified in the function.

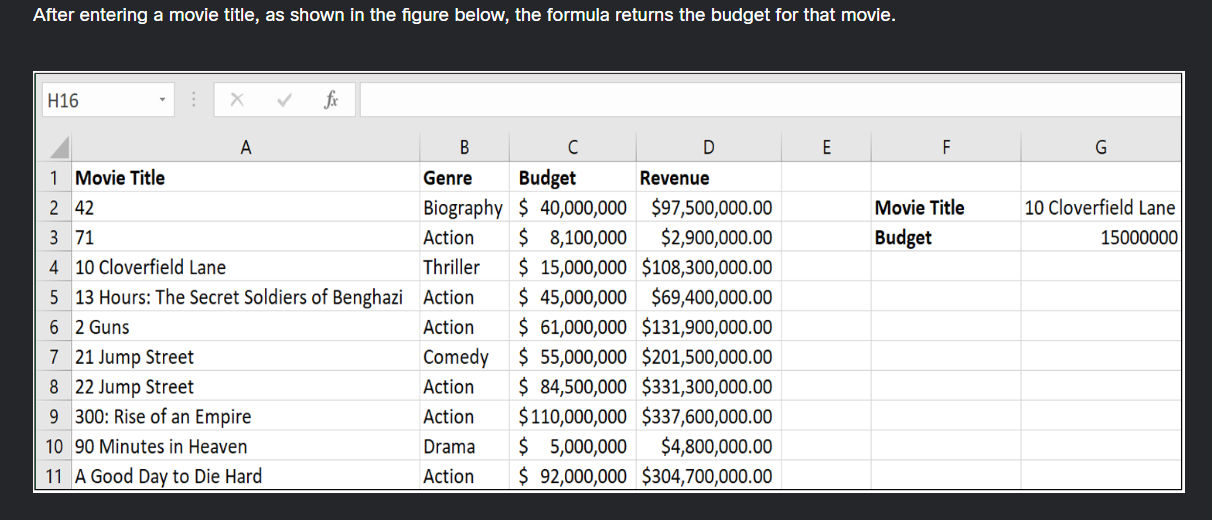
VLOOKUP searches for a value in the leftmost column of a table and, when the value is found, returns information from the same row but in another column.

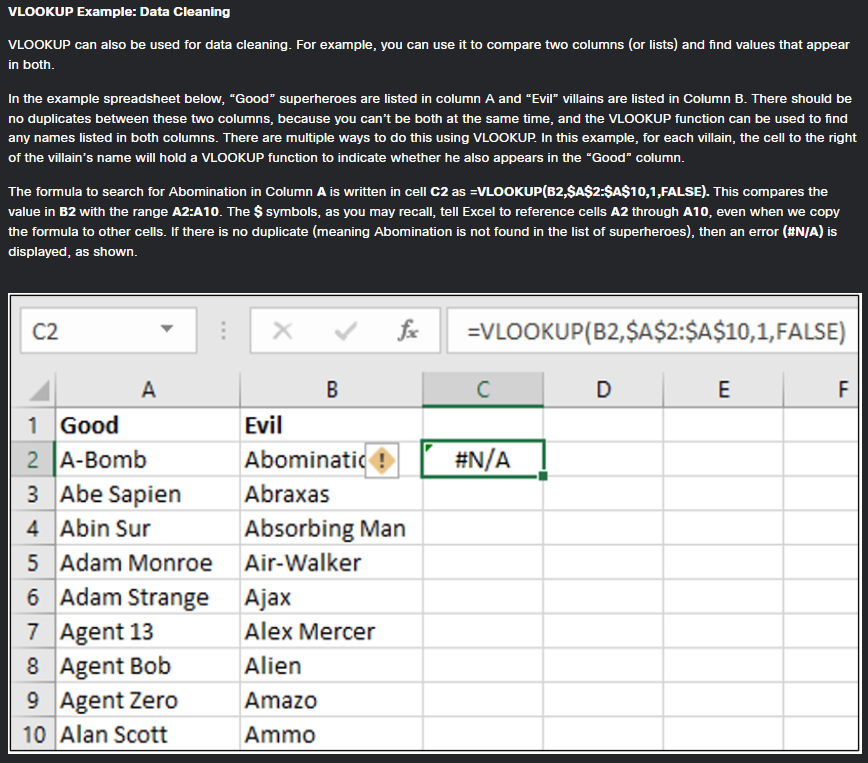
**XLOOKUP, an alternative to VLOOKUP**

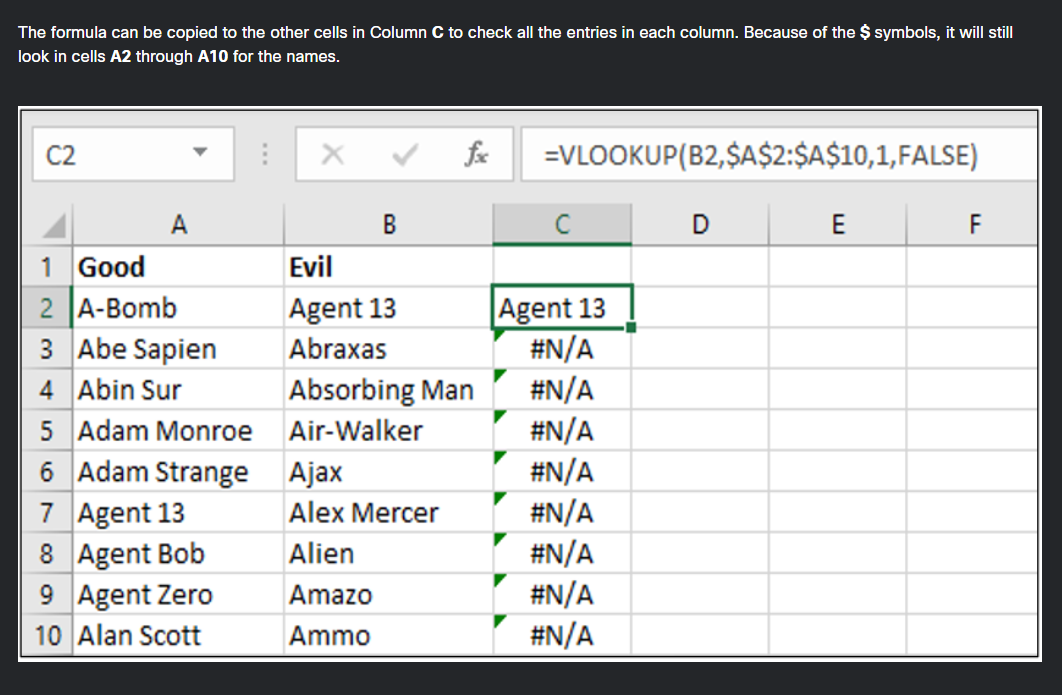
XLOOKUP is a newer lookup function, similar to VLOOKUP, that is not available in all versions of Excel currently in use. With XLOOKUP, you can look in any column (not only the leftmost in a table) for a search term and return a result from the same row. One difference is that XLOOKUP defaults to returning an exact match, whereas VLOOKUP defaults to closest match unless the FALSE keyword is used. In this course, you may use either VLOOKUP or XLOOKUP to obtain the desired results if they are both available in the spreadsheet tool you are using.

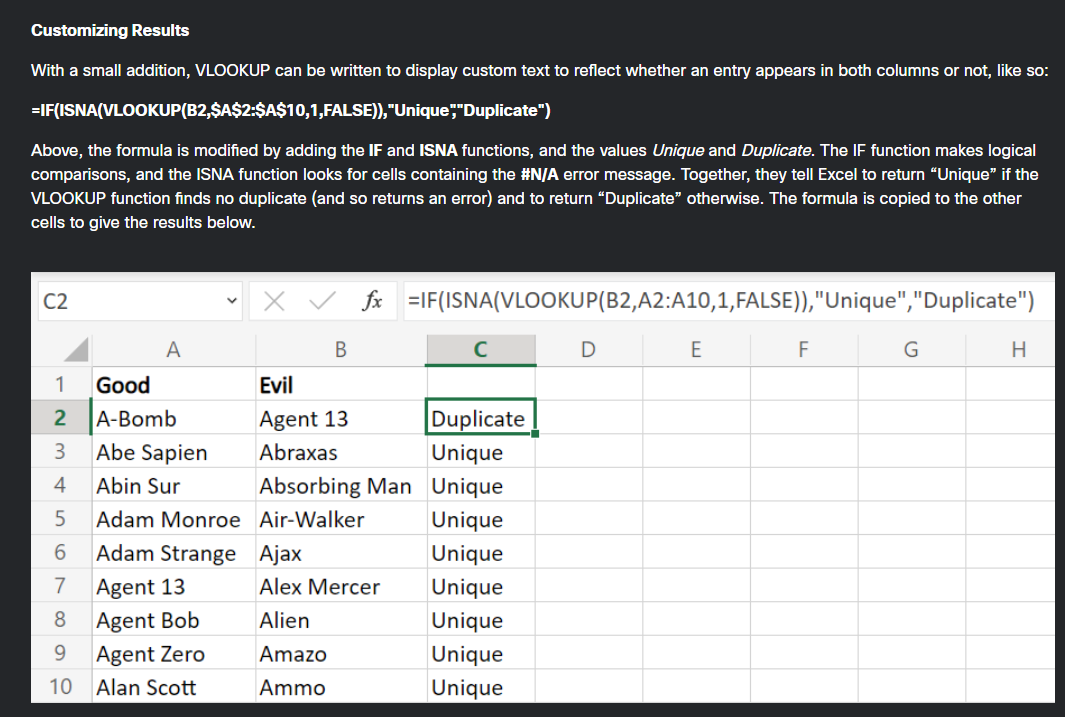
Note: XLOOKUP is not backward compatible, so worksheets using XLOOKUP may not be usable in earlier versions of Excel.



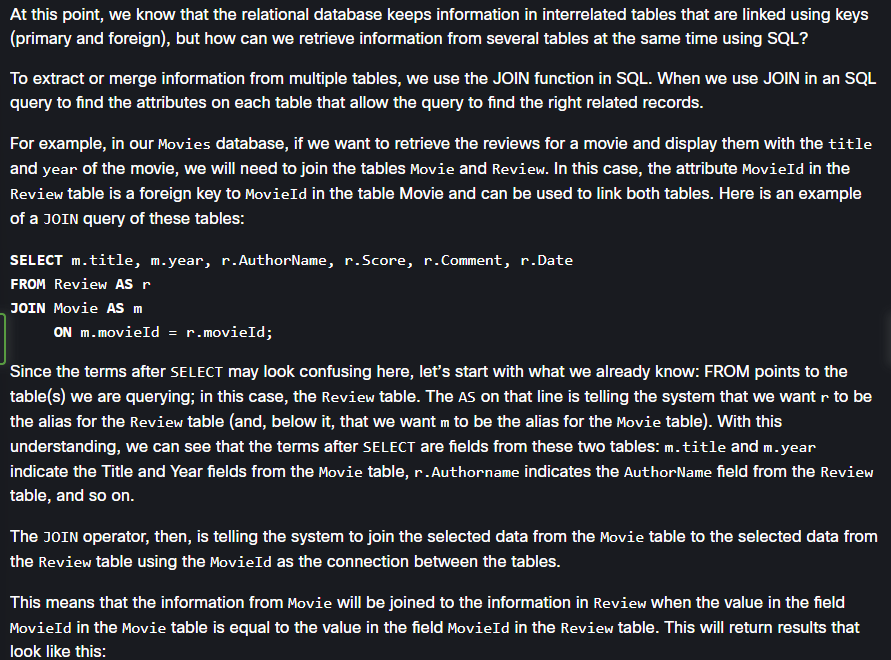


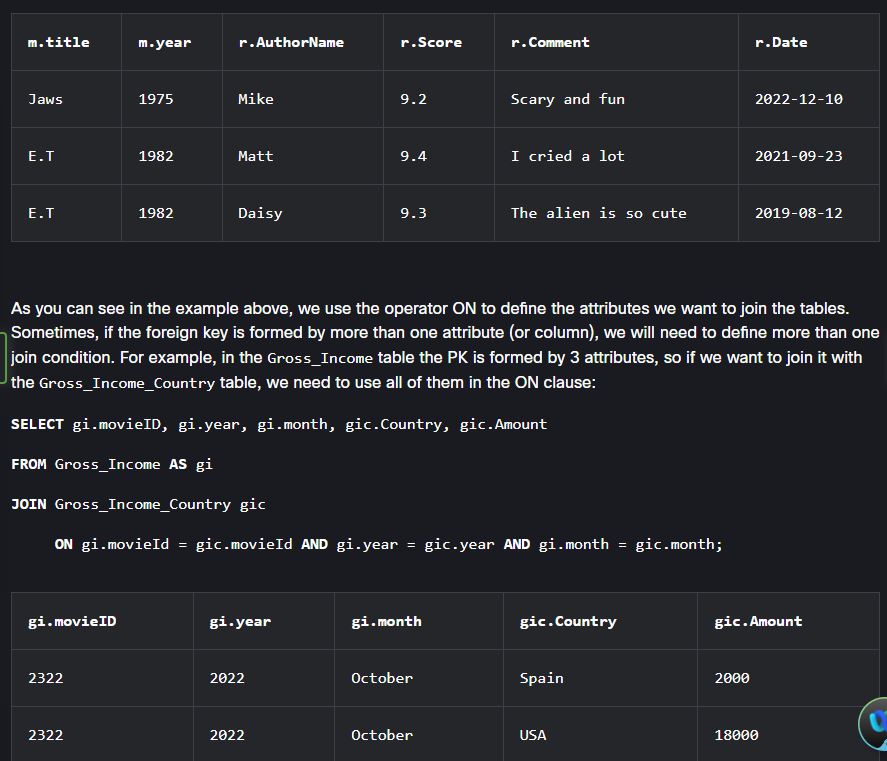


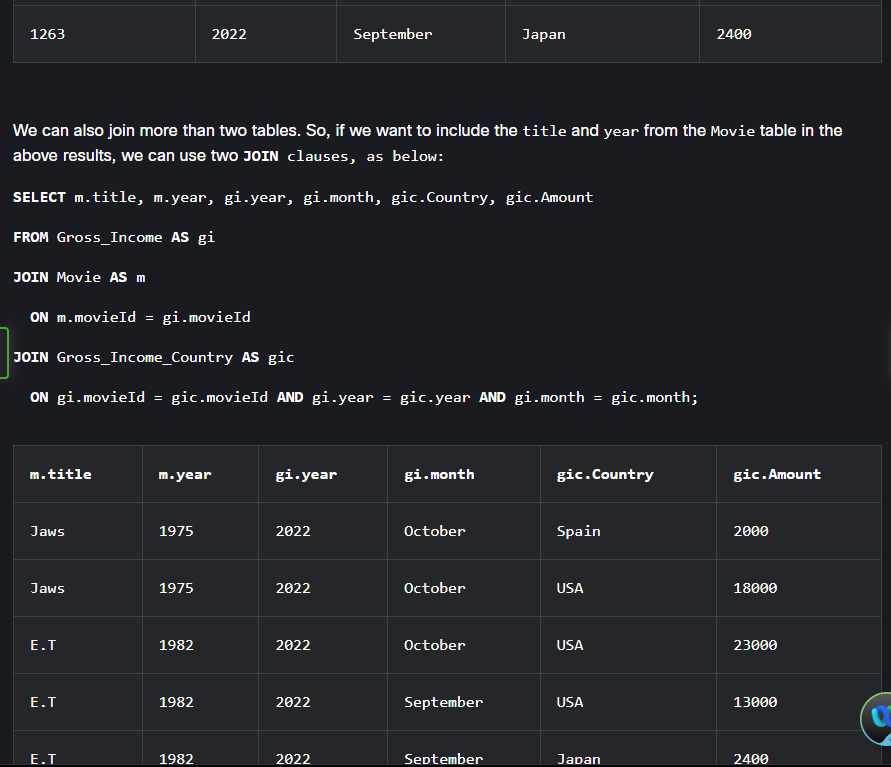




**JOINING TABLES:**

****

****

****

**Types of Joins**

