Spreadsheet Basics

Montag, 30. Juni 2025 22:13

Basic Words
+ Rows (L/R)
+ Columns (U/D)
+ Cells

Attribute
+ characteristic or quality
+ to label a column in a table

Observation
+ all attributes for sth in a row of a data table
+ Row 1 is an observation of Patrick Witczak
+ bc we see his attributes in this row

Formula
+ set of instructions
+ perform specific action
+ using data in spreadsheet

Insert -> Chart
+ creates a chart out of spreadsheet :o
+ many customisations available

End

Datei auswählen Keine ausgewählt

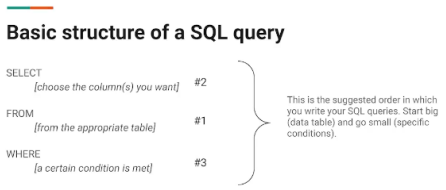
SQL Basics

Montag, 30. Juni 2025 22:13

SQL Basics
+ no matter which database we use
+ SQL works usually the same

Query
+ request for data or information from a database
+ here used Big Query *(Chapter 3)*

First Command
```
SELECT *
FROM movie_data.movies
WHERE Genre__1_ = 'Action'
```
WHERE = filter
a lot of movies appear appeared



Syntax
+ a unique set of guidelines
+ includes required words, symbols, punctations
+ and a proper placement

First Query

Example of a query

Here is how a simple query would appear in BigQuery, a data warehouse on the Google Cloud Platform.

```
SELECT first_name
FROM customer_data.customer_name
WHERE first_name = 'Tony'
```

The above query uses three commands to locate customers with the `first_name`, `'Tony'`:

1. `SELECT` the column named `first_name`
2. `FROM` a table named `customer_name` (in a dataset named `customer_data`)
   (The dataset name is always followed by a dot, and then the table name.)
3. But only return the data `WHERE` the `first_name` is `'Tony'`

Multiple Columns in a query

```
SELECT
    ColumnA,
    ColumnB,
    ColumnC
FROM
    Table where the data lives
WHERE
    Certain condition is met
```

Here is an example of how it would appear in BigQuery:

```
SELECT
    customer_id,
    first_name,
    last_name
FROM
    customer_data.customer_name
WHERE
    first_name = 'Tony'
```

```
SELECT
ColumnA,
ColumnB,
ColumnC
FROM
    Table where the data lives
WHERE
    Condition 1
    AND Condition 2
    AND Condition 3
```

Notice that unlike the `SELECT` command that uses a comma to separate fields / variables / parameters, the `WHERE` command uses the `AND` statement to connect conditions. As you become a more advanced writer of queries, you will make use of other connectors / operators such as `OR` and `NOT`.

Here is a BigQuery example with multiple fields used in a `WHERE` clause:

```
SELECT
    customer_id,
    first_name,
    last_name
FROM
    customer_data.customer_name
WHERE
    customer_id > 0
    AND first_name = 'Tony'
    AND last_name = 'Magnolia'
```

"and" must be used by multiple WHERE
"," is used by multiple SELECT
Multiple FROM did not appear so far

Datei auswählen Keine ausgewählt

database_name"."table_name

Try to use semicolons at the end

```
SELECT field1
FROM table
WHERE field1 = condition;
```

Where Conditions
+ WHERE can help me by putting many criteria
+ one good one is also by choosing rows
+ example - field 1 = 1 "Chavez" *(last name)*

"LIKE" - clause
+ where field 1 must not be identical
+ field 1 LIKE Ch *(search for all names with Ch)*

Select *
+ simply select all columns in the table

How To Do Comments
+ use /* and */
+ use -- *(best practice)*

```
SELECT
    field1 /* this is the last name column */
FROM
    table -- this is the customer data table
WHERE
    field1 LIKE 'Ch%';
```

```
-- This is an important query used later to join with the accounts table
SELECT
    rowkey,  -- key used to join with account_id
Info.date,  -- date is in string format YYYY-MM-DD HH:MM:SS
Info.code  -- e.g., 'pub-###'

FROM  Publishers
```

```
-- Pull basic information from the customer table
SELECT
    customer_id, --main ID used to join with customer_addresss
    first_name, --customer's first name from loyalty program
    last_name --customer's last name
FROM
    customer_data.customer_name
```

Aliases
+ give the things you use more convenient names
+ without actually changing the names
+ for further working

```
SELECT
    my_table_alias.actual_column_name AS my_column_alias
FROM
    actual_table_name AS my_table_alias
```

Real Life Example

```
SELECT
    *
FROM
    Employee
WHERE
    jobCode = 'SFI'
```

A portion of the resulting data returned from the SQL query might look like this:

empID	firstName	lastName	jobCode	salary
0002	Homer	Simpson	SFI	15000
0003	Marge	Simpson	SFI	30000
0034	Bart	Simpson	SFI	25000
0067	Lisa	Simpson	SFI	38000
0088	Ned	Flanders	SFI	42000
0076	Barney	Gumble	SFI	32000

But we only want full time workers

```
SELECT
    *
FROM
    Employee
WHERE
    jobCode <> 'INT'
    AND salary <= 30000;
```

The resulting data from the SQL query might look like the following (interns with the job code **INT** aren't returned):

empID	firstName	lastName	jobCode	salary
0002	Homer	Simpson	SFI	15000
0003	Marge	Simpson	SFI	30000
0034	Bart	Simpson	SFI	25000
0108	Edna	Krabappel	TUL	18000
0099	Moe	Szyslak	ANA	28000

Data Visualisation Basics

Montag, 30. Juni 2025 22:13

Introduction - Data Visualisation

+ Tableau and RStudio are more powerful than Spreadsheet
+ when you work with R, you'll usually use RStudio
+ there are really crazy visualisations
+ there's a lot you can do to build up skills with that

+ don't just see them as facts
+ see them as pictures
+ pictures tell stories

Line charts can track sales over time

Maps can connect sales to locations

Donut charts can show customer segments

Bar charts can compare total visitors and visitors that make a purchase

Datei auswählen Keine ausgewählt

Glossary

Mittwoch, 2. Juli 2025 16:36

Terms and definitions for Course 1, Module 3

Attribute: A characteristic or quality of data used to label a column in a table

Observation: The attributes that describe a piece of data contained in a row of a table

Datei auswählen Keine ausgewählt