



# Introduction à la data et aux bases de données

---

[amigraf.com](https://amigraf.com)



**Introduction aux concepts de base des données :** comprendre la nature et la structure des données

**Microsoft Excel :** l'outil de base pour la manipulation de données

**Conceptualisation d'une base de données :** structurer et relier les données de manière logique

**SQL :** interroger directement les bases de données et extraire des informations précises

# Introduction aux concepts des bases de données

# Qu'est-ce qu'une donnée ?

En informatique, les **données** (**data** en anglais) sont des **représentations d'informations**

Nous pouvons les **stocker**, les **traiter** et les **manipuler**

Elles peuvent prendre des formes ou des **types** variés

## Les données (data)

Je m'appelle Jean Dupont, je suis garagiste et j'ai 25ans

Je m'appelle Marie Dupuis, je suis chauffagiste et j'ai 33ans

Je m'appelle Eric François, je suis contrôleur et j'ai 51ans

## Les données (data)

Je m'appelle **Jean Dupont**, je suis **garagiste** et j'ai **25ans**

Je m'appelle **Marie Dupuis**, je suis **chauffagiste** et j'ai **33ans**

Je m'appelle **Eric François**, je suis **contrôleur** et j'ai **51ans**

# Données structurées, semi-structurées, non structurées

# Les données structurées : organisées et facilement exploitables

Elles peuvent être représentées dans des **tableaux**

Leur **format** est prédéfini

Elles sont **faciles à manipuler**



## Les données (data)

Je m'appelle Jean Dupont, je suis garagiste et j'ai 25ans

Je m'appelle Marie Dupuis, je suis chauffagiste et j'ai 33ans

Je m'appelle Eric François, je suis contrôleur et j'ai 51ans

Prénom	Nom	Métier	Age
Jean	Dupont	garagiste	25
Marie	Dupuis	chauffagiste	33
Eric	François	controleur	51

Quelques exemples de données structurées :

## Le format CSV (comma-separated values)

Le fichier représente un **tableau**

Les données sont séparées par des **virgules**, des points-virgules ou des tabulations


La première ligne du fichier donne le nom des **colonnes**

```
year,position,artist,song,indicativerevenue,us,uk,de,fr,ca,au
"2000","1","Faith Hill","Breathe","24030.051","2","33","-","-","-","-","1"
"2000","2","Santana & The Product G","Maria Maria","23320.084","1","1","-","-","-","-","1"
"2000","3","Joe Thomas","I Wanna Know","21516.777","4","-","-","-","-","-","1"
"2000","4","Aaliyah","Try Again","21099.824","1","5","5","26","-","-","1"
"2000","5","Toni Braxton","He Wasn't Man Enough","21023.066","2","5","-","-","-","-","1"
"2000","6","Rob Thomas & Santana","Smooth","20735.418","1","3","21","-","-","-","1"
"2000","7","Vertical Horizon","Everything You Want","20402.965","1","-","-","-","-","-","1"
"2000","8","Destiny's Child","Say My Name","19489.657","1","3","14","-","-","-","1"
"2000","9","Lonestar","Amazed","19138.169","1","21","91","-","-","-","1"
"2000","10","Matchbox Twenty","Bent","18997.978","1","-","-","-","-","-","1"
"2000","11","Madonna","Music","18983.471","1","1","2","8","1","1","1"
"2000","12","Sisqo","Thong Song","18403.832","3","3","15","15","-","-","1"
"2000","13","Three Doors Down","Kryptonite","18341.509","3","-","-","85","-","-","1"
"2000","14","Destiny's Child","Jumpin' Jumpin'","18020.444","3","5","-","-","-","-","1"
"2000","15","Creed","Higher","17082.223","7","47","-","-","-","-","1"
```

Quelques exemples de données structurées :

## Le tableur

Un programme qui manipule des **feuilles de calcul**, et qui représente également les données en **tableaux**

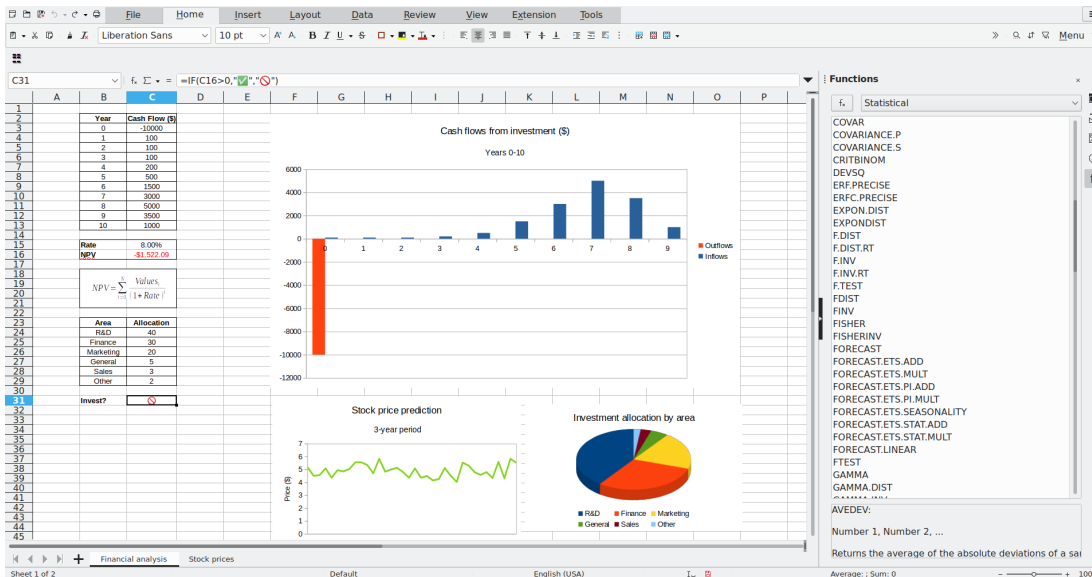


ITEM	NO.	UNIT	COST
MUCK RAKE	43	12.95	556.85
BUZZ CUT	150	6.75	1012.50
TONER	2500	4.95	12487.50
EYE SNUFF	2	4.95	9.90
SUBTOTAL			13155.50
9.75% TAX			1282.66
TOTAL			14438.16

VisiCalc, le premier tableur (1979)

Quelques exemples de données structurées :

## Le tableur



LibreOffice Calc

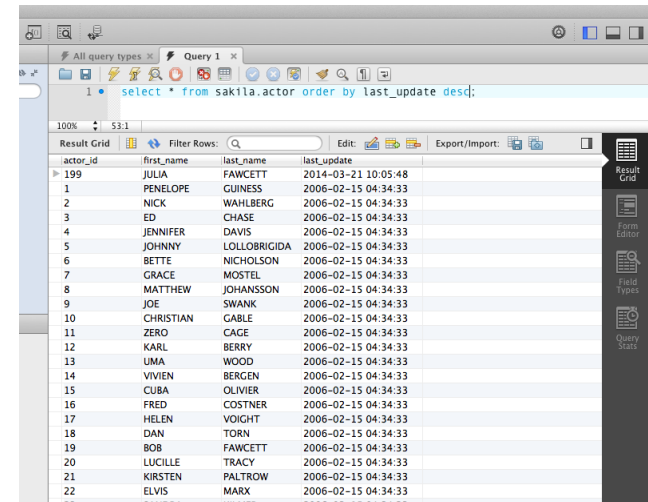
Des fonctionnalités diverses selon les programmes :  
formules et fonctions, graphiques et visualisations, tableau croisé dynamique, macros et scripts...

Quelques exemples de données structurées :

## La base de données relationnelle

Un modèle développé dès le début des années 1970, mais qui demeure l'un des plus utilisés aujourd'hui

Elle représente les données sous forme de **tables**



The screenshot shows the MySQL Workbench interface with a query result grid. The query executed is `select * from sakila.actor order by last_update desc;`. The result grid displays 22 rows of actor data, sorted by last update date in descending order. The columns are actor\_id, first\_name, last\_name, and last\_update.

actor_id	first_name	last_name	last_update
199	JULIA	FAWCETT	2014-03-21 10:05:48
1	PENELOPE	GUINESS	2006-02-15 04:34:33
2	NICK	WAHLBERG	2006-02-15 04:34:33
3	ED	CHASE	2006-02-15 04:34:33
4	JENNIFER	DAVIS	2006-02-15 04:34:33
5	JOHNNY	LOLLOBRIGIDA	2006-02-15 04:34:33
6	BETTE	NICHOLSON	2006-02-15 04:34:33
7	GRACE	MOSTEL	2006-02-15 04:34:33
8	MATTHEW	JOHANSSON	2006-02-15 04:34:33
9	JOE	SWANK	2006-02-15 04:34:33
10	CHRISTIAN	GABLE	2006-02-15 04:34:33
11	ZERO	CAGE	2006-02-15 04:34:33
12	KARL	BERRY	2006-02-15 04:34:33
13	UMA	WOOD	2006-02-15 04:34:33
14	VIVIAN	BERGEN	2006-02-15 04:34:33
15	CUBA	OLIVIER	2006-02-15 04:34:33
16	FRED	COSTNER	2006-02-15 04:34:33
17	HELEN	VOIGHT	2006-02-15 04:34:33
18	DAN	TORN	2006-02-15 04:34:33
19	BOB	FAWCETT	2006-02-15 04:34:33
20	LUCILLE	TRACY	2006-02-15 04:34:33
21	KIRSTEN	PALTROW	2006-02-15 04:34:33
22	ELVIS	MARX	2006-02-15 04:34:33

Un aperçu de base de données dans le logiciel MySQL Workbench

# Les données non structurées : riches en informations, mais complexes à traiter

Abondantes dans le **Big Data**, elles sont faciles à accumuler

Exemples : e-mails, photos, fichiers audio...

Elles ne peuvent **pas** être correctement représentées sous la forme d'un tableau

**Éric François** <eric.francois@mail... 09:30 (il y a 6 minutes) ☆ ↩ ⋮

À Marie Dupuis ▼

Joyeux anniversaire Marie ! J'espère que tu fêtes ça dignement, 33 ans c'est pas rien ! Comment tu vas ?

Ça fait un bail, j'espère te recroiser bientôt. T'as des nouvelles de Jean ? Je me demandais si il était toujours garagiste.

Si ça t'intéresse, je connais une pâtisserie qui fait des fraisiers super bons, celle qui est rue Fragaria. Pas d'anniversaire sans gâteau !

À bientôt,

Éric



**Éric François** <eric.francois@mail... 09:30 (il y a 6 minutes) ☆ ↩ ⋮

À Marie Dupuis ▼

Joyeux anniversaire Marie ! J'espère que tu fêtes ça dignement, 33 ans c'est pas rien ! Comment tu vas ?

Ça fait un bail, j'espère te recroiser bientôt. T'as des nouvelles de Jean ? Je me demandais si il était toujours garagiste.

Si ça t'intéresse, je connais une pâtisserie qui fait des fraisières super bons, celle qui est rue Fragaria. Pas d'anniversaire sans gâteau !

À bientôt,

Éric

Leur **traitement** est plus complexe : data science, intelligence artificielle, machine learning...

Nom	Prénom	Métier	Âge	Va peut-être s'acheter un fraisier
François	Éric	?	33	oui
Dupuis	Marie	?	?	?
?	Jean	Garagiste ?	?	?

# Excel

**Un exemple de stockage et de traitement de  
données structurées**

# Excel : qu'est ce que c'est ?



Microsoft Excel est un logiciel **tableur** de la suite bureautique Microsoft Office

Il manipule des **feuilles de calcul**, contenues dans des **classeurs**

Il utilise (entre autres) des fichiers **.xlsx**

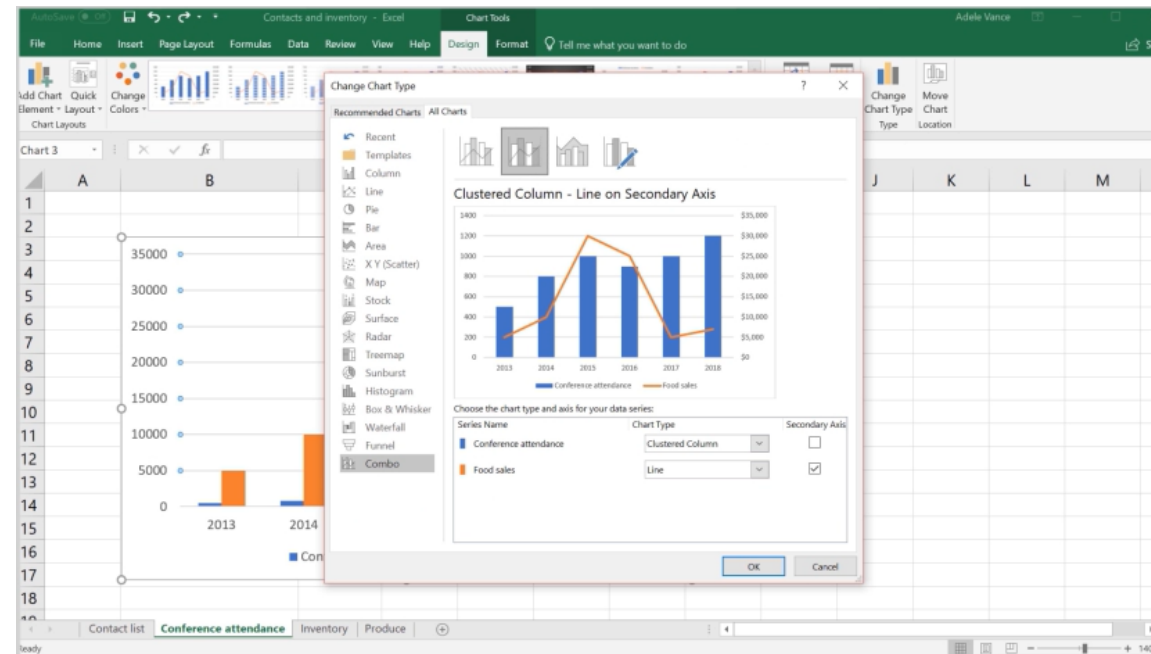
Excel intègre des fonctionnalités :

- de **calcul numérique**
- d'**analyse de données**
- de **représentation graphique**
- et même de **programmation**, avec les macros et Visual Basic for Applications (VBA)

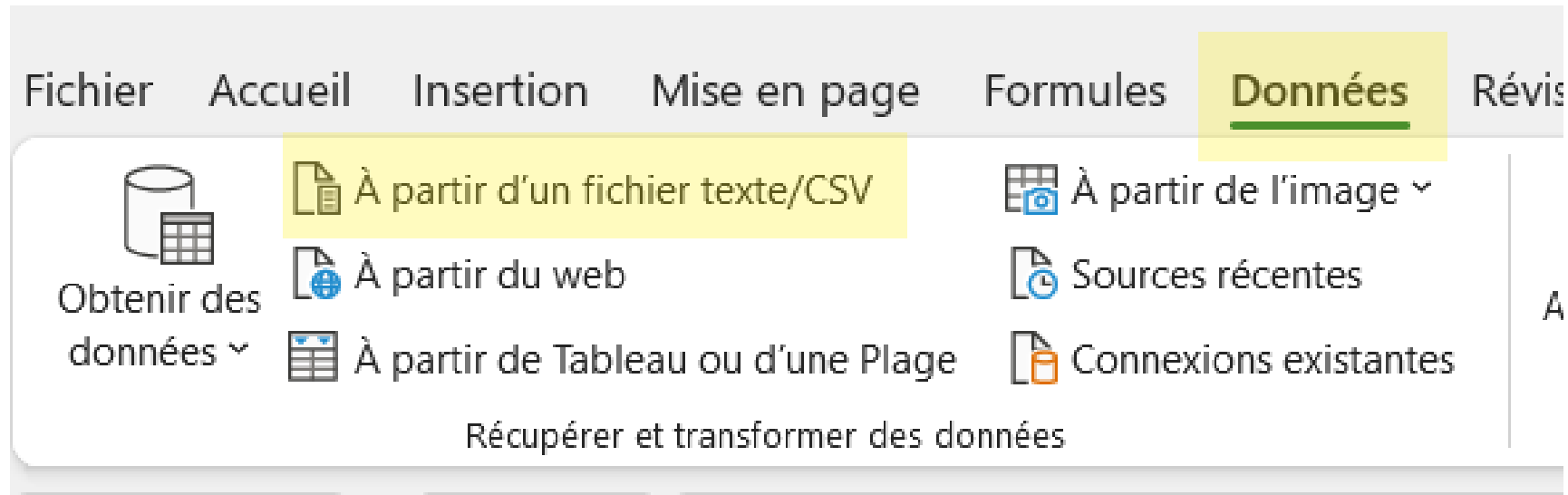


De nombreux plugins et programmes **tiers** fonctionnent avec Excel

Il est intuitif et **facile d'utilisation** pour les non-spécialistes : recommandations, outils de visualisation clairs, de nombreuses fonctions pré-définies, tableau croisé dynamique...



# Import de fichiers .csv



# Qu'est-ce qu'une cellule ?

Les **colonnes** sont nommées par des **lettres**

Les **lignes** sont nommées par des **chiffres**

	A	B	C
1	A1	B1	C1
2	A2	B2	C2
3	A3	B3	C3
4	A4	B4	C4
5	A5	B5	C5

Une cellule est l'**intersection** entre une colonne et une ligne

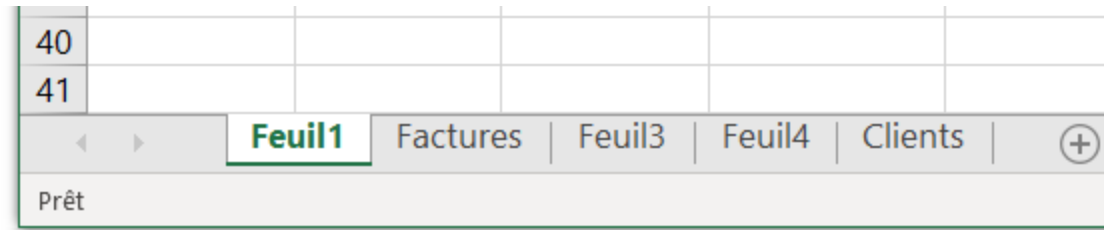
Son **adresse**, sa **référence**, sera donc composée d'une lettre et d'un chiffre : **A5**, **B3**...

On peut y écrire une **donnée**, mais aussi une **formule** ou une **fonction**

Lorsqu'on veut faire **référence** à une cellule qui se trouve sur **une autre feuille** de calcul :

**=Feuille!Cellule**

Par exemple :



**=Achats!B5**

fait référence à la cellule **B5** sur une feuille qui s'appelle **Achats**



# Opérations arithmétiques de base : +, -, \*, /

L'addition, la soustraction, la multiplication et la division

On peut les effectuer sur des nombres, ou sur des **références** de cellules

**=10+5**

**=A1\*B3**

**=(C4-5)\*2**

Pour signifier qu'on va écrire une opération ou une fonction, on commence par le **signe égal =**

## La recopie incrémentée

élève n°3	27	jeudi	2 mars 2010
élève n°5	26	samedi	3 août 2010
élève n°7	25	lundi	4 janvier 2011
élève n°9	24	mercredi	7 juin 2011
élève n°11	23	vendredi	8 novembre 2011
élève n°13	22	dimanche	10 avril 2012
élève n°15	21	mardi	11 septembre 2012
élève n°17	20	jeudi	12 février 2013

Pour **verrouiller** la référence de la cellule, on utilise le caractère **dollar \$**

	A	B	C
1	Prix HT	Prix TTC	
2	200	240	=A2*(1+\$B\$8)
3	470	564	=A3*(1+\$B\$8)
4	140	168	=A4*(1+\$B\$8)
5	100	120	=A5*(1+\$B\$8)
6	54	64,8	=A6*(1+\$B\$8)
7			
8	Taux de TVA :	20%	
9			

# Quelques fonctions statistiques simples

## **SOMME(), MIN(), MAX()**

Permettent d'obtenir la somme, la valeur minimale et la valeur maximale

Exemple : **=MAX(B4:E4)** donne la plus grande valeur de la ligne B4:E4

Les arguments peuvent être des **valeurs**, des références à des **cellules**, ou à des **plages** de cellules

Exemple : **=SOMME(B12:B15; 3; C8)** additionne toutes les cellules de B12 à B15, le chiffre 3, et la cellule C8

# MOYENNE(), MEDIANE(), ECARTYPE.STANDARD()

Permettent d'obtenir la moyenne, la médiane et l'écart-type

Exemple : **=ECARTYPE.STANDARD(B2:B9)**

calcule l'écart-type des valeurs qui se trouvent dans la colonne B2:B9

# Attention à la syntaxe !

**Plusieurs éléments** sont séparés par des **points-virgules ;**  
**Une plage** de cellules est indiquée par **deux points :**

**=MIN(A2;A6)**

renvoie la plus petite valeur entre **deux cellules**, A2 et A6

**=MIN(A2:A6)**

renvoie la plus petite valeur **parmi toutes les cellules** qui se trouvent entre A2 et A6 compris

	A	B	C
1			
2	450	=MIN(A2;A6)	
3	27	MIN(nombre1; [nombre2]; [nombre3]; ..	
4	95		
5	12	résultat : 27	
6	27		
7			

	A	B	C
1			
2	450	=MIN(A2:A6)	
3	27	MIN(nombre1; [nombre2]; ...)	
4	95		
5	12	résultat : 12	
6	27		
7			



On peut **combina** les fonctions : faire des moyennes de sommes, trouver la valeur minimale entre deux moyennes, etc.

Par exemple,

**=MIN(SOMME(B1:B7);SOMME(C1:C7))**

Renvoie **la plus petite valeur** entre **la somme de la colonne B** et **la somme de la colonne C**

## Générer des nombres aléatoires : **ALEA()**, **ALEA.ENTRE.BORNES()**

=**ALEA()** génère un nombre supérieur ou égal à 0 et inférieur à 1

Exemple : **ENT(ALEA()\*100)** : génère un nombre entier supérieur ou égal à 0 et inférieur à 100

## Générer des nombres aléatoires : **ALEA()**, **ALEA.ENTRE.BORNES()**

**=ALEA.ENTRE.BORNES(1; 50)** : génère un nombre entre 1 et 50

⚠ Les nombres sont générés à chaque fois qu'on édite le tableau !

## Les dates : un type particulier de données

Excel stocke les dates sous une forme particulière pour qu'elles soient utilisables dans des calculs (numéros de série)

Lorsqu'on importe des données, Excel essaie de reconnaître les dates et les convertit

Très souvent cela fonctionne, parfois il y a besoin d'un traitement supplémentaire

# Fonctions de date

**=DATE(annee ; mois ; jour)** crée une date  
inversement, **=ANNEE(date)** extrait l'année d'une date

D2	x	✓	fx	=DATE(C2;A2;B2)	Fonction DATE
	A	B	C	D	E
1	Mois	Jour	Année	Date combinée	
2	3	14	2012	14/03/2012	
3					
4					

Combine 3 valeurs en 1

**=MAINTENANT()**

# **TP : Analyse de la qualité de l'air sur une année**

# Fonctions de texte

## GAUCHE(), DROITE()

**GAUCHE(texte; nb)** et **DROITE(texte; nb)**

prennent en entrée une chaîne de **texte**, et facultativement un **nombre de caractères**

Si le nombre de caractères n'est pas spécifié, la valeur par défaut est de **1**

Exemples :

**=GAUCHE("Pastèque"; 3)** retourne "Pas"

**=GAUCHE(A3)** retourne la première lettre de la chaîne qui se trouve dans la cellule A3 (un seul caractère)

**=DROITE(B7; 4)** retourne les quatre dernières lettres de la cellule B7



# STXT()

**STXT(texte; position; nb)**

retourne un **nombre** donné de caractères à partir d'une **position** donnée

Par exemple, **=STXT(C1; 3; 5)** renvoie 5 caractères de la cellule C1, en commençant par le 3e caractère

## NBCAR()

**NBCAR(texte)** retourne la taille de la chaîne (son nombre de caractères)

Par exemple, **=NBCAR("Coucou")** donne 6

## MAJUSCULE(), MINUSCULE()

Convertissent un texte en majuscules ou en minuscules

**=MINUSCULE("bONJouR")** a pour résultat "bonjour"

➔ Permet de normaliser les données

# SUBSTITUE()

**Remplace** des caractères dans une chaîne de texte

**=SUBSTITUE(chaine; ancien\_texte; nouveau\_texte; position)**

**chaine** : du texte ou une référence à une cellule

**ancien\_texte** sera remplacé par **nouveau\_texte**

**position** (facultatif) : si spécifié, on remplace uniquement cette occurrence, sinon toutes les occurrences sont remplacées

## CHERCHE()

**=CHERCHE(texte\_recherche; ou\_chercher;)**

Retourne la position d'un caractère/d'une chaîne dans une autre chaîne : par exemple, **=CHERCHE("n","bonjour")** renvoie 3

**TROUVE()** : même chose que CHERCHE(), mais sensible à la casse

# Fonctions de recherche : **RECHERCHEV()**, **RECHERCHEH()**, **INDEX()**, **EQUIV()**

## **RECHERCHEV()**

Cherche une valeur dans la première colonne du tableau

Renvoie la valeur d'une cellule qui se situe sur la même ligne que la valeur recherchée

**=RECHERCHEV(recherche; tableau; colonne; type)**

**=RECHERCHEV(C3;C2:F7;2;FAUX)**

**Le premier argument** : la valeur de recherche  
une référence (C3), ou une valeur (du texte, un nombre...)

**Le deuxième argument**

La **plage** de cellules (C2:F7) dans laquelle on va chercher le résultat

**Le troisième argument**

Le numéro de la colonne **dans cette plage de cellules** où l'on va  
chercher le résultat

**=RECHERCHEV(C3;C2:F7;2;FAUX)**

**Le dernier argument** : VRAI/FAUX (facultatif, par défaut = **VRAI**)

Est-ce qu'on peut renvoyer une correspondance **approximative** ou non

**FAUX** ou **0** : on renvoie uniquement une **correspondance exacte**

**VRAI** ou **1** (ou vide) : on cherche la valeur **la plus proche**



## =RECHERCHEV(recherche; tableau; colonne; type)

	A	B	C	D
1	<b>Prénom</b>	<b>Nom</b>	<b>Courriel</b>	<b>Téléphone</b>
2	Roger	Dupont	<a href="mailto:roger.dupont@mail.com">roger.dupont@mail.com</a>	01 02 03 04 05
3	Bob	Martin	<a href="mailto:bob.martin@mail.com">bob.martin@mail.com</a>	06 12 34 56 78
4	Micheline	Durand	<a href="mailto:micheline.durand@mail.com">micheline.durand@mail.com</a>	01 23 45 67 89
5	Joséphine	Bernard	<a href="mailto:josephine.bernard@mail.com">josephine.bernard@mail.com</a>	07 89 10 11 12
6				
7				
8		Mail de Bob : <a href="mailto:bob.martin@mail.com">bob.martin@mail.com</a>		
9			=RECHERCHEV("bob";A2:D5;3;FAUX)	
10				

	A	B	C	D	E	
1	Prénom	Roger	Bob	Micheline	Joséphine	
2	Nom	Dupont	Martin	Durand	Bernard	
3	Courriel	roger.dupont@r	bob.martin@m	micheline.durar	josephine.berna	
4	Téléphone	01 02 03 04 05	06 12 34 56 78	01 23 45 67 89	07 89 10 11 12	
5						
6		Mail de Bob : bob.martin@mail.com				
7			=RECHERCHEH("bob";B1:E4;3;FAUX)			

## =RECHERCHEH()

fait la même chose que **RECHERCHEV()**, mais dans l'autre sens :  
cherche dans la première ligne d'un tableau, et renvoie la valeur  
d'une cellule qui se situe dans la même colonne

Limites de **RECHERCHEV()** :

Pour l'utiliser, il faut déjà savoir dans quelle **colonne** se trouvera le **résultat**

La valeur de recherche ne peut être que dans la **première colonne**

On ne peut donc chercher que de **gauche à droite**, les résultats ne peuvent pas être à gauche de valeur de recherche

**à demain !**