



UNIVERSITÉ PARIS 1

PANTHÉON SORBONNE



UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE



1.3 Traiter des données

Sommaire

- Donnée et connaissance
- L'open data
- Hub de données
- Data Lab, Data Lake
- Data Warehouse, Data Mining
- Le Machine Learning
- Google données
- Références "mixtes"
- Conversion texte-colonne
- Fonctions texte
- Afficher un caractère de formule

Donnée et Connaissance

- Depuis plusieurs années la **donnée** (information) est toujours plus au cœur de l'entreprise. Dans le but d'automatiser toujours plus leurs traitements, de gagner en productivité et d'être toujours plus proches de leur marché ou de leur écosystème, les organisations se sont lancées dans une collecte toujours plus massive de données.
- Plus les entreprises réussissent à tirer parti de ces données, plus elles augmentent cette collecte pour être toujours plus justes, plus précises et plus rapides en termes de compréhension de leur environnement et pertinentes en termes de prise de décision.
- Cette course à la performance et à l'efficacité s'est logiquement accompagnée d'une diversification des provenances et des sources de données.
- A travers ces trois étapes précitées, la **connaissance** apparaît comme le stade représentant le plus de valeur. Il faut bien prendre en compte que cela a été rendu possible du fait de l'existence et de la mise à disposition des données préalablement stockées.

Donnée et Connaissance

- Les données peuvent être stockées dans une **base de données** (accessibles par des requêtes **SQL** pour les **BD relationnelles**) ou dans des **fichiers** (accessibles par des **programmes** écrits avec un **langage de programmation** comme Python, Java, C++, C, ...).
- Ces BD et fichiers peuvent être stockés dans l'entreprise ou à l'extérieur de l'entreprise. Le problème posé est qui a droit ou pas à accéder à la donnée de l'entreprise.
- « **Le cloud** » fait référence aux **serveurs** accessibles via Internet, ainsi qu'aux logiciels et bases de données qui s'exécutent sur ces serveurs. Les serveurs cloud sont situés dans des **datacenters** partout dans le monde. En utilisant le cloud computing, les utilisateurs et les entreprises n'ont pas à gérer eux-mêmes les serveurs physiques ni à exécuter des applications logicielles sur leurs propres machines. Le cloud offre un grand nombre de services aux entreprises et ces services sont facturés aux entreprises à l'usage. Les entreprises peuvent déléguer la gestion et le fonctionnement de leur système d'information aux fournisseurs de services cloud (Microsoft, Google, Amazon, ...) ce qui leur permet de réaliser des économies sur les salaires, pas d'achat de machines, ni de logiciel, ... et de se concentrer sur leur cœur de métier.

Donnée et Connaissance

- Pour en savoir plus :
- <https://www.cloudflare.com/fr-fr/learning/cloud>
- <https://www.bing.com/videos/search?q=cloud+azure+youtube&docid=607990605591836424&mid=3FFADED8602D672294DD3FFADED8602D672294DD&view=detail&FORM=>
- <https://consent.youtube.com/m?continue=https://www.youtube.com/watch%3Fv%3DRZdesXYQuYY&gl=FR&m=0&pc=yt&uxe=23983171&hl=fr&src=1VIRE/what-is-the-cloud/>

L'open data

- L'Open data rend les **données accessibles et utilisables par tous** et disponibles en ligne
- Pour les collectivités locales et les organismes publiques, l'open data signifie publier sur une **plateforme ouverte** des informations : statistiques, cartographie, horaires...
 - Exemple : pour une mairie, les horaires des bus
- La mise à disposition des données publiques est une obligation légale
 - Les données sensibles et personnelles sont exclues de la démarche open data
- Les développeurs peuvent utiliser ces données pour développer des applications.

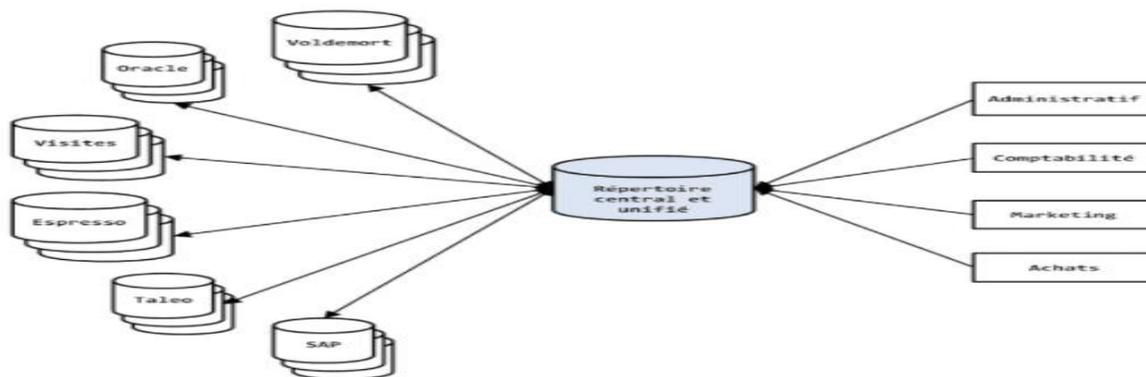
L'open data

- Données gouvernementales <https://www.data.gouv.fr/fr/>
- Données sur la ville de Paris <https://indicateurs-parisdata.opendatasoft.com/pages/home/>
- Données SNCF <https://data.sncf.com/pages/accueil/>
- Données énergie ENEDIS <https://www.enedis.fr/open-data>
- CNIL <https://www.cnil.fr/fr/definition/open-data>

- Données sur la santé <https://www.health-data-hub.fr/>

Hub de données

- Le **Data Hub** a été mis sur pied au même titre que le Data Lab, le Data Lake ou le Data Warehouse , comme stratégie pour aider les entreprises à uniformiser toutes leurs sources de données et les rendre facilement exploitable.
- L'objectif du Data Hub est de “**casser les silots**” c'est-à-dire unifier les données dans un lieu physique ou un répertoire de stockage physique pour en faciliter l'accès et l'utilisation.
- un Data hub, encore appelé « *hub des données* », est une plateforme de stockage de données virtuelles provenant de diverses sources. Il déplace et intègre physiquement des données multi-structurées et les stocke dans un répertoire unique.



Datalab, Data Lake

- Le **Datalab** est une solution marketing pour améliorer sa technique de fidélisation clients. Il prend en compte le comportement de la clientèle et décrit précisément les habitudes de consommation pour un produit spécifique.
- Un **Data Lake** est un référentiel de données permettant de stocker une très large quantité de données brutes dans le format natif pour une durée indéterminée. Cette méthode de stockage permet de faciliter la cohabitation entre les différents schémas et formes structurelles de données, généralement des **blobs** (binary large object: données codées en binaire) ou des fichiers.



Data Warehouse, Data Mining

Le *Data Warehouse*, ou entrepôt de données, est une base de données dédiée au stockage de l'ensemble des données utilisées dans le cadre de la prise de décision et de l'analyse décisionnelle.

Le Data Warehouse est alimenté en données depuis les bases de production grâce notamment aux outils d'**ETL** Extract Transform Load.

Ensuite, les utilisateurs, analystes et décideurs, accèdent aux données Collectées, historisées, et mises en forme pour étudier des cas précis de réflexion. Ils construisent des modèles d'étude et de prospective pour limiter la part d'incertitude lors du processus de décision.



Data Warehouse, Data Mining

- Définition 1 : Le data-mining est un processus de découverte de règles, relations, corrélations et/ou dépendances à travers une grande quantité de données, grâce à des méthodes statistiques, mathématiques et de reconnaissances de formes.
- Définition 2 : Le data-mining est un processus d'extractions automatique d'informations prédictives à partir de grandes bases de données.
- Exemple: analyser l'opinion des usagers sur les produits d'une entreprise à travers les commentaires sur les réseaux sociaux et les blogs.



Big Data et Machine Learning

- Le **Big Data** désigne à la fois la production de données massives et le développement de technologies capables de les traiter afin d'en extraire des corrélations ou du sens. C'est dans les années 1990 que le terme Big Data prend sa signification actuelle d'un défi technologique à relever pour analyser de grands ensembles de données, d'abord scientifiques, mais de plus en plus souvent collectés au quotidien par divers moyens techniques (téléphone portable ou autres objets connectés dans l'Internet des Objets, réseaux sociaux, capteurs, caméras...).
- Le **Machine Learning (ML)** utilise des probabilités statistiques pour donner aux ordinateurs la capacité d'apprendre par eux-mêmes. Pour ce faire, le data scientist utilise des algorithmes programmés le plus souvent en langage Python ou R. Le ML utilise des données massives (du big data) pour la mise en œuvre des algorithmes.
- Le ML fait partie intégrante de **l'Intelligence Artificielle (IA)**. C'est l'ensemble des théories et des techniques mises en œuvre en vue de réaliser des machines capables de simuler l'intelligence humaine

Google données



Google données

- Pour créer un graphique
 - (1) Sélectionnez un type de graphique au moyen du sélecteur de type de graphique.
 - (2) Sélectionnez les statistiques à afficher dans votre graphique en cliquant sur le nom des statistiques dans le volet des données.
 - (3) Choisissez les entités à comparer en cliquant sur les cases à cocher du volet "Comparer par".
 - (4) Choisissez les valeurs à filtrer en sélectionnant les entités dans le volet de filtre.
 - (5) Cliquez sur le bouton "Lecture" (graphique à barres, carte ou graphique à bulles) pour animer les barres ou les bulles dans le graphique au fil du temps.

Références "mixtes"

L'utilisation du symbole \$ dans Excel

- Pour ne pas avoir à ressaisir des formules dans toutes les cellules d'un tableau, il est possible d'utiliser une formule qui va imposer à Excel d'adapter la formule de départ à l'ensemble des cellules concernées.
- On va alors figer la formule grâce au symbole \$. On parle alors de référence mixte.
- Une référence mixte est une référence qui n'est fixée que sur une partie de la référence : soit la ligne, soit la colonne. Elle comprend soit une colonne absolue et une ligne relative, soit une ligne absolue et une colonne relative.
 - Une référence de colonne absolue s'affiche sous la forme \$A1 (bloque la colonne), à l'inverse, une référence de ligne absolue s'affiche sous la forme A\$1 (bloque la ligne).

Références "mixtes" : Exemple

Dans cet exemple nous appliquons le même prix unitaire à chaque produit.

=C4*\$B\$1													
	A	B	C	D	E	F	G	H	I	J	K	L	
1	Prix	20											
2			Volumes des ventes					Chiffre d'affaire					
3			Marc	Sophie	Adrien	Totaux		Marc	Sophie	Adrien	Totaux		
4		Panier	13	45	14	72		Panier	260	900	280	1440	
5		Lampe	51	12	35	98		Lampe	1020	240	700	1960	
6		Table	43	20	17	80		Table	860	400	340	1600	
7		Totaux	107	77	66	250		Totaux	2140	1540	1320	5000	

Pour caculer le chiffre d'affaire, il suffit de se placer dans une cellule et de réaliser un premier calcul à partir d'une formule classique. Ex :

=C4*B1

Pour faire de même avec d'autres cellules, il suffit alors de figer la référence du prix avec le symbole \$ pour appliquer correctement ce prix à toutes les cellules concernées, comme illustré dans l'exemple suivant :

=C4*\$B\$1

Il ensuite possible de tirer la formule en allant vers la droite puis vers le bas.

A NOTER : Ici nous avons figé la colonne et la ligne mais il est possible de figer la formule seulement sur une ligne :

Ex : =C4*B\$1

Ou à l'inverse seulement sur une colonne :

Ex : =C4*\$B1

Conversion texte-colonne

Dans un tableur il est possible de fractionner du texte en plusieurs colonnes à partir de l'onglet "Convertir" lorsque vous souhaitez importer des données dans un tableur ou séparer des noms et des prénoms dans deux colonnes distinctes.

- Sélectionnez la cellule ou la colonne contenant le texte à fractionner ;
- Sélectionnez **Données > Convertir** ;
- Dans l'**Assistant Conversion**, sélectionnez **Délimité > Suivant** ;
- Sélectionnez les **Séparateurs** de vos données. Par exemple, **Virgule** et **Espace**. Un aperçu de vos données apparaît dans la fenêtre **Aperçu des données** ;
- Sélectionnez **Suivant** ;
- Sélectionnez la **Destination** des fractions de données sur votre feuille de calcul ;
- Sélectionnez **Terminer**.

Fonctions texte

Il est fréquent, dans Excel, de manipuler le contenu d'une cellule pour :

- **Extraire** une sous-chaînes ;
- **Fusionner** le contenu de plusieurs cellules (on parle de concaténation) ;
- **Supprimer** une partie des informations.

Pour **extraire à l'intérieur d'une chaîne**, on utilise la **fonction STXT**.

Dans cet exemple, pour extraire le département nous allons écrire :

=STXT(A2;6;2)

Arguments de la fonction

STXT

Texte A2 = "2740418047158"

No_départ 6 = 6

No_car 2 = 2

Renvoie un nombre déterminé de caractères d'une chaîne de texte à partir de la position que vous indiquez.

No_car indique le nombre de caractères à extraire du texte.

Résultat = 18

Aide sur cette fonction

OK Annuler

Indiquer depuis quelle cellule l'extraction doit être réalisée.

Fonctions texte Gauche/droite

La fonction **GAUCHE** récupère les n caractères qui commencent une chaîne de caractères, à l'inverse la fonction **DROITE** récupère ceux qui terminent la chaîne.

Arguments de la fonction

GAUCHE

Texte A2 = "2740418047158"

No_car 1 = 1

Résultat = 2

Extrait le(s) premier(s) caractère(s) à l'extrême gauche d'une chaîne de texte.

No_car indique le nombre de caractères que la fonction GAUCHE doit renvoyer; 1 si omis.

Pour sélectionner une fonction : clic gauche sur l'icône fonction > catégorie TEXTE > puis choisir la fonction GAUCHE, DROITE ou STXT.

La formule concerne ici la première ligne, c'est-à-dire la cellule A2.

Emplacement du genre dans le numéro de sécurité sociale. Il s'agit du premier caractère c'est donc ce chiffre que l'on indique va reporter dans le tableau.

	A	B	C	D
1	N° de sécurité sociale	Département	Numéro étudiant	Genre
2	2 74 04 18 047 158	56	112 335	2
3	1 87 11 91 031 138	75	115 451	1
4	2 55 06 28 345 285	91	113 798	2
5				

Fonctions texte (STXT)

	A	B	C	D	E	F
1	N° de sécurité sociale	Numéro étudiant	Genre	Département		
2	2 74 04 18 047 158 *	112 335	2	18		
3	1 87 11 19 031 138	115 451	1	19		
4	2 55 06 28 345 285	113 798	2	28		
5						

* Pour modifier format > clic droit "Format de cellule"

Le département correspond aux deux chiffres entourés ici. Il s'agit donc des caractères 6 et 7 de la chaîne de texte.
No_départ à l'emplacement du premier chiffre concerné,
No_car indique le nombre de caractères à extraire du texte soit 2 dans cet exemple. Les chiffres doivent correctement apparaitre entre les guillemets.

Arguments de la fonction

STXT

Texte	A2	= "2740418047158"
No_départ	6	= 6
No_car	2	= 2
		= "18"

Renvoie un nombre déterminé de caractères d'une chaîne de texte à partir de la position que vous indiquez.

No_car indique le nombre de caractères à extraire du texte.

Résultat = 18

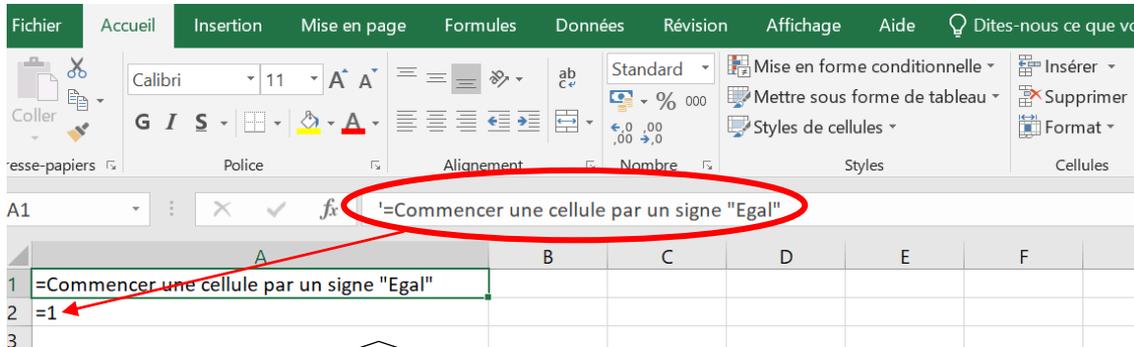
[Aide sur cette fonction](#)

OK Annuler

Afficher un caractère de formule

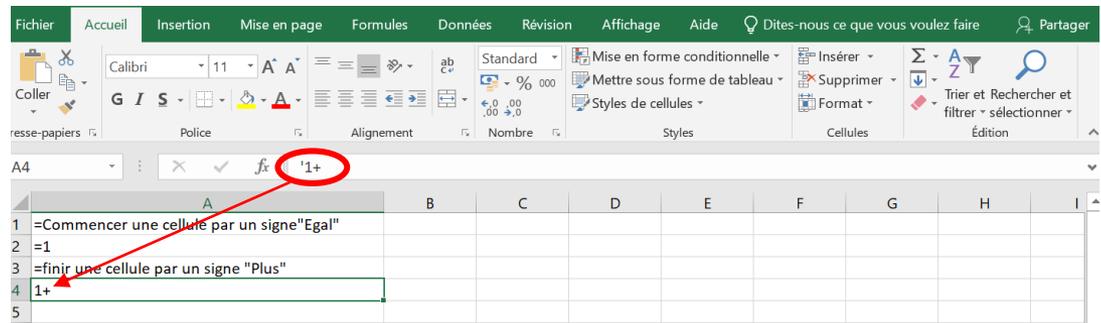
Il est possible d'afficher un caractère de formule (=, -, +, etc.) dans une cellule sans pour autant que le logiciel ne l'interprète comme telle. Ces caractères apparaissent alors dans le résultat.

Dans Excel il suffit de saisir une **apostrophe** dans la barre de formule :



L'apostrophe peut être également utilisé devant n'importe quels autres signes * / - pour écrire du texte.

Excel propose de saisir une **apostrophe** , visible **uniquement** dans la barre de formule et non dans le résultat, puis vos **données**.



DSIUN – Service des usages numériques

- **Conception – Réalisation**
 - Catherine Loire – Mélanie Mauvoisin - Service des usages numériques - Université Paris 1 Panthéon-Sorbonne

Version du support : 1.0
Juin 2021



[Licence Creative Commons :](#)
[Paternité - Pas d'Utilisation Commerciale](#)
[Partage des Conditions Initiales à l'Identique](#)



UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE
